UNIVERSITY OF INFORMATION TECHNOLOGY

FACULTY OF INFORMATION SYSTEMS

Forecasting the closing price trends of crypto Using statistical models and machine learning

- VO HUNG KHANG - 20521440
- VUONG THI YEN NHI - 20521721
- HUYNH KHA TU - 20522096

FORECASTING THE CLOSING PRICE TRENDS OF CRYPTO USING STATISTICAL MODELS AND MACHINE LEARNING

1st Vo Hung Khang
STAT3013.N11.CTTT
University of Information Technology
Ho Chi Minh city
20521440@gm.uit.edu.vn

2nd Vuong Thi Yen Nhi
STAT3013.N11.CTTT
University of Information Technology
Ho Chi Minh city
20521721@gm.uit.edu.vn

3rd Huynh Kha Tu
STAT3013.N11.CTTT
University of Information Technology
Ho Chi Minh city
20522096@gm.uit.edu.vn

Abstract — Crypto has been a long-lasting and potential investment field attracting much investment in this field every year. In particular, favorite cryptos such as Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) have attracted many investments in recent years. The volatility of crypto is very unpredictable that makes investor have difficulty in investing this field. The goal of this paper is determine what accuracy the direction of crypto price in USD can be forecasted. This study will use models like Linear regression [1], Non-linear regression [2], ARIMA [3], SARIMA [4], LSTM [5] and Prophet [6] to forecasting the close price of three cryptos BTC, ETH, LTC. The comparison results will be based on three evaluation parameters: RMSE, MAE, and MAPE.

Keywords — *Bitcoin prediction, Ethereum prediction, Tether prediction, machine learning, linear regression, non-linear regression, ARIMA, SARIMA, LSTM, Prophet.*

I. INTRODUCTION

After cryptocurrency (CTC) has been used, cryptocurrencies have become a global phenomenon that is a digital medium of exchange. Bitcoin is the first and best-known cryptocurrency, while Ethereum or Tether is one of cryptocurrency become popular too. Cryptocurrency price prediction is one of the trending areas among researchers. This paper makes an attempt to apply statistics models as linear, nonlinear regression, ARIMA and machine learning like LSTM, Prophet for predicting close (closing) price of cryptocurrency (Bitcoin, Ethereum and Tether). After that, we will compare these model and find the model with the best accuracy to predict future close price.

II. RELATED WORK

Reaz Chowdhury et al. [7] conducted Cryptocurrencies Index 30 (cci30) forecasts from January 1st 2015 to January 1st 2017 throught two variants of Gradient Boosted Trees, Neural Network, Ensemble Learning Method and K-NN models. In this case, the K-NN model didn't worked effectively unlike other models.

Sean McNally et al. [8] conducted Coindesk Bitcoin Price forecasts from the 19th of August 2013 until the 19th of July 2016 throught two variants of RNN and LSTM models. The result shows that the LSTM outperformed the RNN marginally, but not significantly.

V. Derbentsev et al. [9] conducted Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP) forecasts from the 19th of August 2013 until the 19th of July 2016 throught RNN and LSTM. This study adjusted the units and epochs and used the MAPE. The result shows that the out of sample accuracy of short-term forecasting daily prices obtained by RNN and LSTM in terms of MAPE for three of the most capitalized cryptocurrencies (BTC, ETH, and XRP) was within 0.92-2.61%.

V. Derbentsev et al. [10] conducted BTC-USD, ETH-USD, BNB-USD, LTC-USD, XLM-USD, and DOGE-USD forecasts from January 1, 2018 to December 31, 2021 throught ADAB, GBM, XGB, DFNN, GRUs, CNNs. This study adjusted the units and epochs and used the NSE, EVS, t-test, MAPE . The result shows that the CNN, GRU, and DFNN models produce high correlation coefficients, low RMSE, and standard deviations from the measured observations, compared to tree-based models that did not work effectively for some cryptocurrencies.

Mohammed Mudassir et al. [11] conducted Bitcoin (BTC) forecast for the period of April 1, 2013, to December 31, 2019 through ANN, SANN, SVM and LSTM models. The comparison results will be based on three evaluation parameters: RMSE, MAE, and MAPE. The result shows that all the developed models are satisfactory and have good performance.

Mohammed Mudassir et al. [12] conducted Coinmarketcap (BTC) forecast for the period of March 15, 2020 to June 28, 2020 throught LSTM, ARIMA, SARIMAX and SVM models. The result shows that the machine learning method SVM and the time series forecasting method SARIMAX achieved most successful results in terms of estimating ‘‘Bitcoin Trend Detection problem for the crisis period".

Negar Maleki et al. [13] conducted Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC) and Zcash forecast from 1st April 2018 to 31st March 2019 throught MA, ARIMA, SVM, Linear Regression, Lasso Regression algorithm, RFR. The result shows that Zcash had the most trend's similarity toward Bitcoin by executing the Lasso Regression algorithm as an ML's method.

Yan Li and Wei Dai [14] conducted Kraken Bitcoin forecast from 30 December 2016 to 31 August 2018 throught CNN, LSTM, CNN-LSTM hybrid. The results show that CNN-LSTM hybrid neural network performs well in Bitcoin forecast.

Alvin Ho et al. [15] conducted Bitcoin (BTC) forecast from 29th August 2017 to 9th August 2020 throught Linear Regression and LSTM. The results show that the best accuracy rate is shown in LSTM.

R. K. Jana et al. [16] conducted Bitcoin (BTC) forecast January 10, 2013 to February 23, 2019 throught MODWT, LSTM, SVR, PRI, and DE models. The results show that the best accuracy rate is shown in LSTM. This research proposes a novel DE-based regression framework for forecasting one day ahead price of Bitcoin. The result shows that the utilization of the DE for optimizing the learning algorithms' performance assists in augmenting the forecast quality that helps us forecast accurater.

III. METHODS

A. *Data Collection*

We using dataset of three cryptocurrency include: Bitcoin, Ethereum and Tether from 26/12/2017 to 26/12/2022 (during 5 years).

B. *Linear Regression*

This technique is used to identify the relationship between dependent and independent variables and is leveraged to predict future outcomes. When we use only one dependent and one independent variable then it is called the simple linear regression. As the number of independent and dependent variable increase, it is then referred to as multi-linear regression. [17] Linear regression is both a statistical algorithm and a machine learning algorithm. Regression has the following general expression:

$$f(x) = a + bx$$

With:

- f(x): is the output.
- x: is the input (indepent variable).
- a: is constant.
- b: is the coefficient of linear equation.

C. *Non-Linear Regression*

Nonlinear regression is a mathematical model that fits an equation to certain data using a generated line. As is the case with a linear regression that uses a straight-line equation (such as Ỵ= c + m x), nonlinear regression shows association using a curve, making it nonlinear in the parameter.

The nonlinear regression model when at least one parameter is a nonlinear function:

$$y = a + \sum_{i=1}^{m} a_i f_i(x_1, x_2, \dots, x_k)$$

Formula:

A simple nonlinear regression model is expressed as follows:

$$Y = f(X, \beta) + \epsilon$$

Where:

- X is a vector of P predictors
- β is a vector of k parameters
- F (-) is the known regression function
- ϵ is the error term

D. *ARIMA*

ARIMA stands for Auto Regressive Integrated Moving Average. It is a well-known forecasting model in financial and data science time series application [18]. The auto-regressive moving average (ARMA) models are used in stationary cryto data only, this model contains three combination models which are AR (p), MA (q) and I (d).

We have the autoregression model:

$$A(p) = Yt = c1 + \alpha 1Yt\text{-}1 + \alpha 2Yt\text{-}\ 2 + \cdots + \alpha pYt\text{-}p + \varepsilon t$$

$$= c1 + \sum pi{=}1\ \alpha i\ Yt\text{-}i + \varepsilon t$$

And the moving average model:

$$M(q) = Yt = c1 + \beta 1\varepsilon t\text{-}1 + \beta 2\varepsilon t\text{-}\ 2 + \cdots + \beta q\varepsilon t\text{-}q$$

$$= c1 + \sum qi{=}1\ \beta i\ \varepsilon t\text{-}i$$

From that, ARIMA model has the form

$$ARIMA\ (p, d, q) = \mathrm{Yt}\ = \mathrm{AR}(p) + I(d) - MA(q)$$

E. *SARIMA*

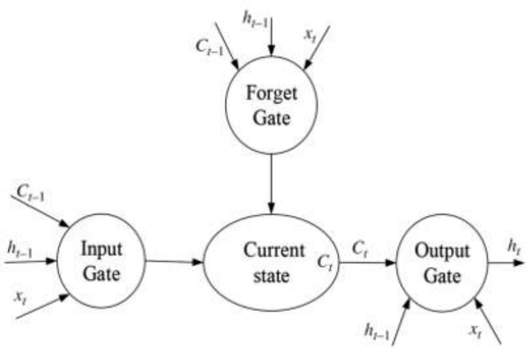
This is where SARIMA (Seasonal Autoregressive Integrated Moving Average) models come in. As an extension of the ARIMA method, the SARIMA model not only captures regular difference, autoregressive, and moving average components as the ARIMA model does but also handles seasonal behaviour of the time series. 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. [20]

In general, the ARIMA Seasonal model is notated as follows:

$$\mathrm{SARIMA}(p,d,q)(P, D, Q)s$$

F. *LSTM (Long Short-Term Memory)*

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN), which is the general term of a series of neural networks capable of processing sequential data. LSTM is a special network structure with three “gate” structures (shown in Fig. 1). Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM's network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate.[21]



- Forget gate: f_t = σ (U_f*x_t + W_f*h_{t-1}+b_f)
- Input gate: i_t = σ (U_i*x_t + W_i*h_{t-1}+b_i)
- Output gate: o_t = σ (U_o*x_t + W_o*h_{t-1}+b_o)

The input gate determines how much new information is added to the neuron state.

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t$$

The output gate is used to control how many current neural unites state are fltered and how many controlling units state are fltered

$$h_t = O_t * \tanh(C_t)$$

G. Prophet

Prophet is an open source software that is available in Python and R for forecasting time series data. PROPHET is published by Facebook's Core Data Science team. It depends on a contribution model where non-linear trends are fit with weekly and yearly seasonality and plus holidays. PROPHET is strong to missing data, capturing the shifts in the trend and large outliers. In addition, it gets a reasonable estimate of the mixed data without spending manual effort [22]

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t$$

- g(t): trend (non-period changes)
- s(t): seasonality (periodic changes)
- h(t): holiday effect
- ε_t : error term

IV. EVALUATION METHODOLOGY

In this research, predictive models are evaluated according to three criteria: MAE, MAPE, and RMSE.

- Mean Absolute Error – MAE

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|$$

- Mean Absolute Percentage Error – MAPE

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|$$

- Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Includes:

- n: is sample size of dataset.
- y_i : is the actual value at time t.
- $\bar{y}_i$: is the mean value at time t.
- y_i : is the predicted value of time t.

V. ANALYSIS

A. Visualization

In this study, we use three CTC, Bitcoin, Ethereum and Litecoin during 5 years from 26/12/2017 to 26/12/2022.

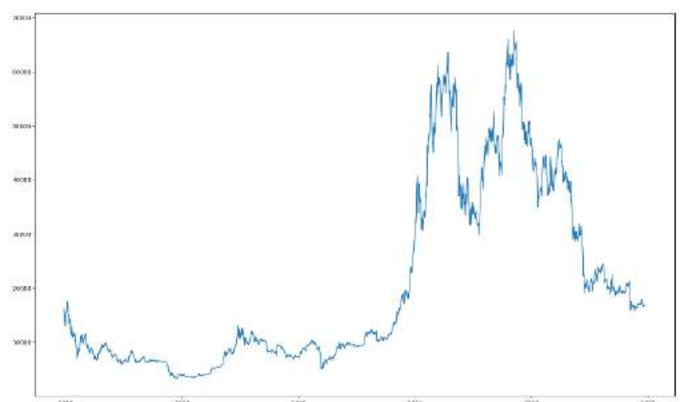


Figure 1: The figure of the closing price of Bitcoin

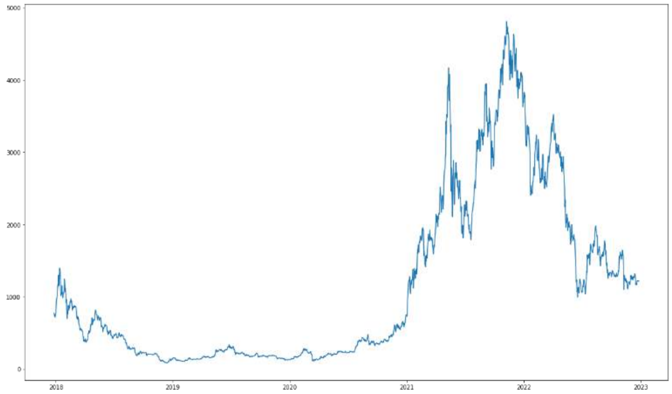


Figure 2: The figure of the closing price of Ethereum

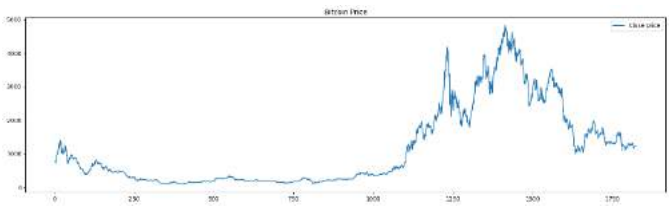


Figure 3: The figure of the closing price of Litecoin

B. Splitting data

At the same time, we split the data sets into 70% training data - 30% testing data, 80% training data - 20% testing data and 90% training data – 10% testing data.

VI. RESULT

Table 1: Evaluation of single models with closing of the Bitcoin's price.

Model		RMSE	MAPE	MAE
		Closing	Closing	Closing
LR	10 - 0	390844.9	3556.6%	390475.9
NLR	10 – 0	256812.3	2666.5%	164125.8
ARIMA	9 – 1	2522.73	11.22%	2084
	8 – 2	24101.43	99.59%	21479
	7 – 3	14272.92	38.83%	12369
SARIMA	9 – 1	4380.12	19.6%	4198.29
	8 – 2	26534.9	60.48%	26237.7
	7 – 3	16618	49.35%	16554.1
LTSM	9 – 1	**875.75**	**3.04%**	**600.42**
	8 – 2	2176.3	6.23%	1683.46
	7 – 3	5939.11	14.29%	4855.31

Given Bitcoin's price, the LSTM model is more suitable than rest in terms of most indicators such as MAE, RMSE and MAPE.

Table 2: Evaluation of single models with closing of the Ethereum's price.

Model		RMSE	MAPE	MAE
		Closing	Closing	Closing
LR	10 - 0	27361.98	7540.3%	27335.66
NLR	10 - 0	15483.4	6542.5%	169812.1
ARIMA	9 – 1	**297.5**	14.34%	2975
	8 – 2	2214.49	133.76%	2064
	7 – 3	1231.76	36.75%	985.22
SARIMA	9 – 1	867.46	40.77%	853.97
	8 – 2	2126.8	62.1%	2076.76
	7 – 3	1204.19	37.15%	973.09
LTSM	9 – 1	106.26	**5.53%**	**77.3**
	8 – 2	364.59	14.51%	281.36
	7 – 3	441.70	12.77%	325.4
PROPHET	9 – 1	735	26.6%	412
	8 – 2	467	23.6%	272
	7 – 3	326	24.1%	169

Beside Bitcoin's price, Ethereum's price is still suitable for LSTM model with the ratio 90% of training data and 10% of testing data (expressed throgh MAPE and MAE), while ARIMA model is good expressed through RMSE.

Table 3: Evaluation of single models with closing of the Litecoin's price.

Model		RMSE	MAPE	MAE
		Closing	Closing	Closing
LR	10 - 0	128789.1	2023.8%	349854
NLR	10 – 0	245317.4	2589.4%	175642
ARIMA	9 – 1	19.64	31.13%	17.87
	8 – 2	76.56	111%	71.5
	7 – 3	58.61	67%	50.31
SARIMA	9 – 1	16.25	30.05%	15.86
	8 – 2	68.99	48.76%	67.6
	7 – 3	60.77	45.78%	60.07
LTSM	9 – 1	**3.04**	**3.61%**	**2.16**
	8 – 2	5.13	5.26%	3.92
	7 – 3	13.74	6.92%	8.29
PROPHET	9 – 1	52	31.25%	34
	8 – 2	45	28.55%	30
	7 – 3	43	30.1%	28

The LSTM model with the ratio of 90% training data and 10% test data gave suitable for Litecoin's price.

- Visualize predict price:

A. Bitcoin

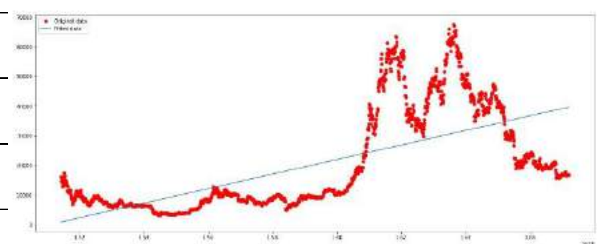


Figure 4: The figure of Linear Model of Bitcoin price result

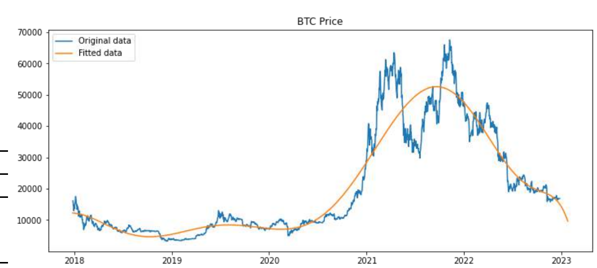


Figure 5: The figure of NonLinear Model of Bitcoin price result

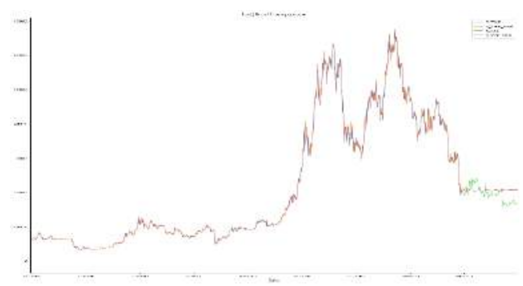


Figure 6: The figure of ARIMA Model of Bitcoin price result (9 – 1)

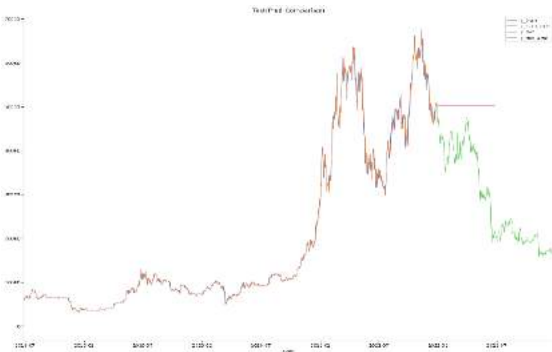


Figure 7: The figure of ARIMA Model of Bitcoin price result (8 – 2)

Figure 8: The figure of ARIMA Model of Bitcoin price result (7 – 3)

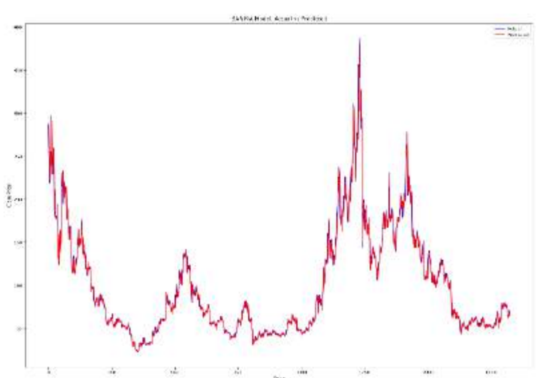


Figure 9: The figure of SARIMA Model of Bitcoin price result (9 – 1)

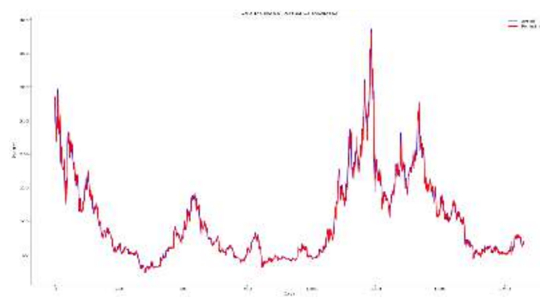


Figure 10: The figure of SARIMA Model of Bitcoin price result (8 – 2)

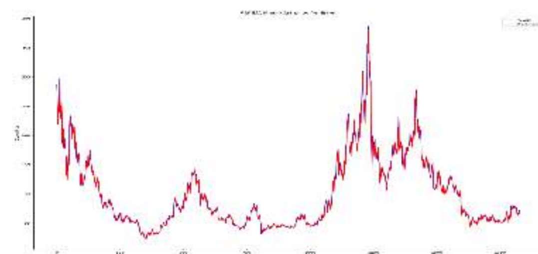


Figure 11: The figure of SARIMA Model of Bitcoin price result (7 – 3)

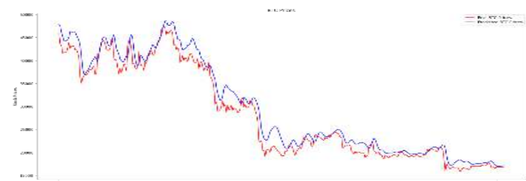


Figure 12: The figure of LSTM Model of Bitcoin price result (8 – 2)

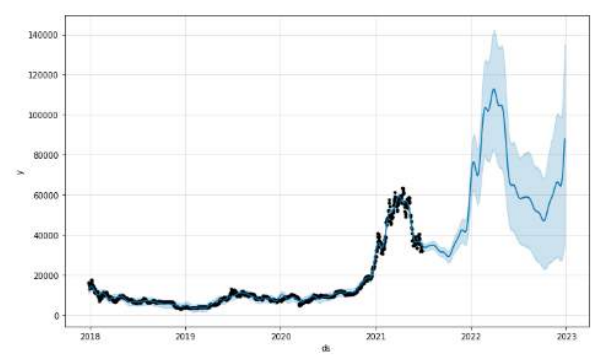


Figure 13: The figure of Prophet Model of Bitcoin price result (7 – 3)

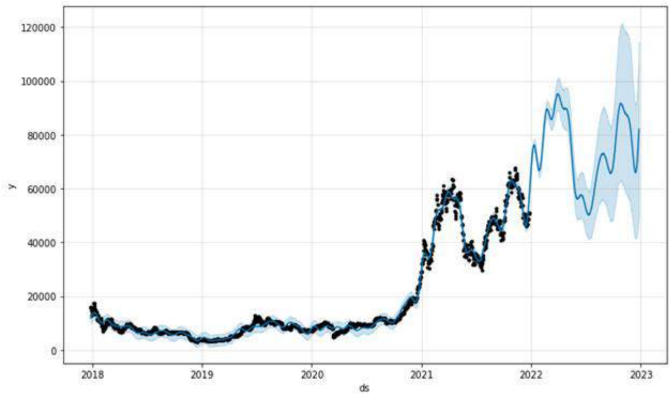


Figure 14: The figure of Prophet Model of Bitcoin price result (8 – 2)

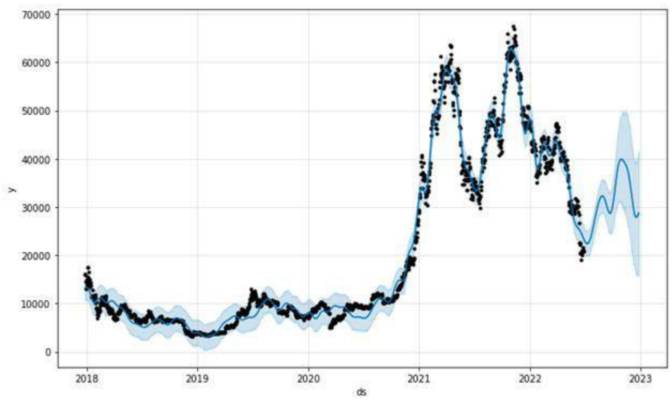


Figure 15: The figure of Prophet Model of Bitcoin price result (9 – 1)

B. Ethereum

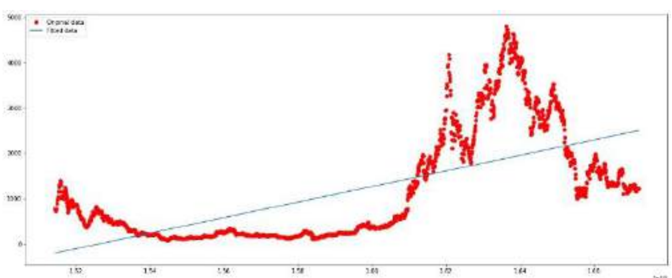


Figure 16: The figure of Linear Model of Ethereum price result

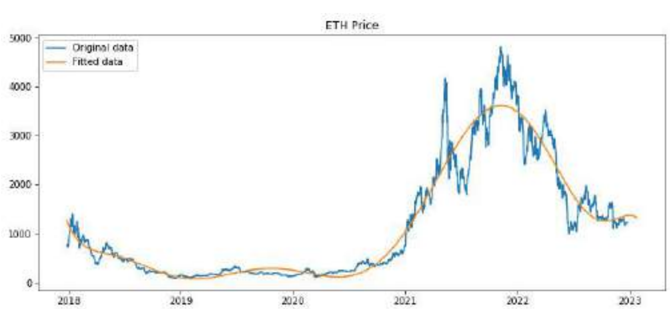


Figure 17: The figure of NonLinear Model of Ethereum price result

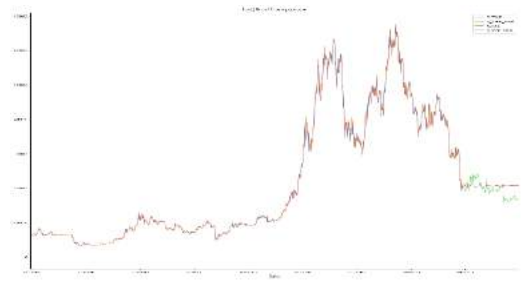


Figure 18: The figure of ARIMA Model of Ethereum price result (9 – 1)

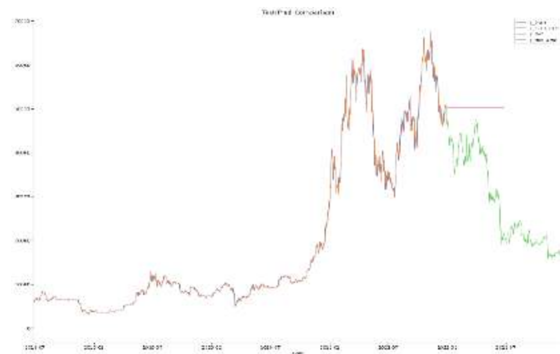


Figure 19: The figure of ARIMA Model of Ethereum price result (8 – 2)

Figure 20: The figure of ARIMA Model of Ethereum price result (7 – 3)

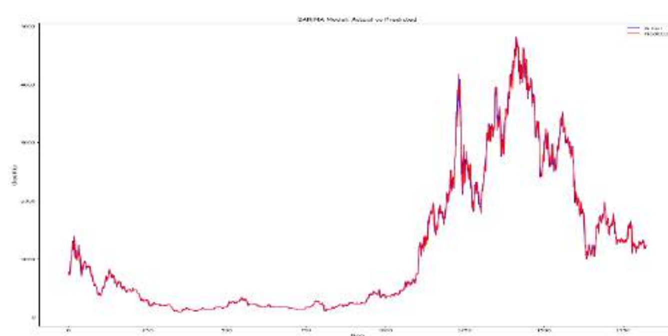


Figure 21: The figure of SARIMA Model of Ethereum price result (9 – 1)

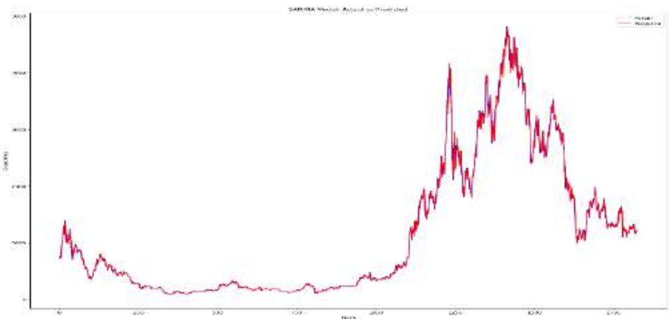


Figure 22: The figure of SARIMA Model of Ethereum price result (8 – 2)

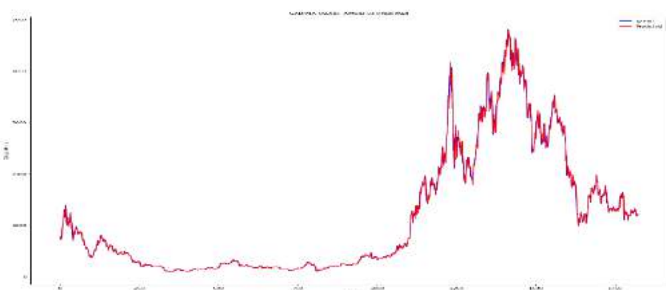


Figure 23: The figure of SARIMA Model of Ethereum price result (7 – 3)

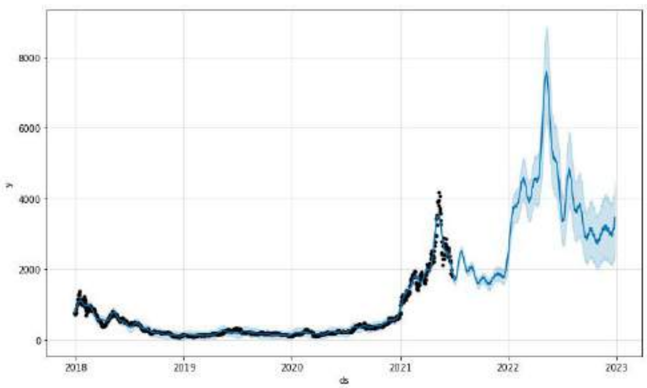


Figure 24: The figure of Prophet Model of Ethereum price result (7 – 3)

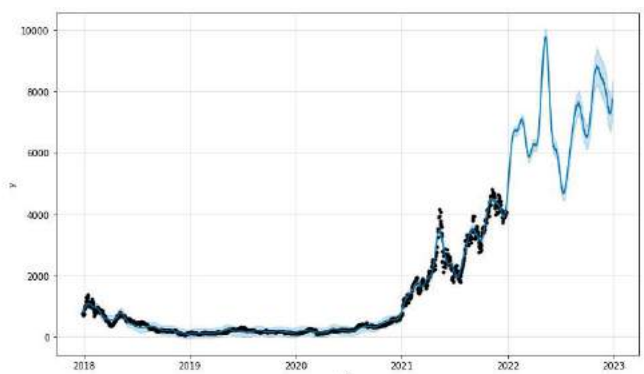


Figure 25: The figure of Prophet Model of Ethereum price result (8 – 2)

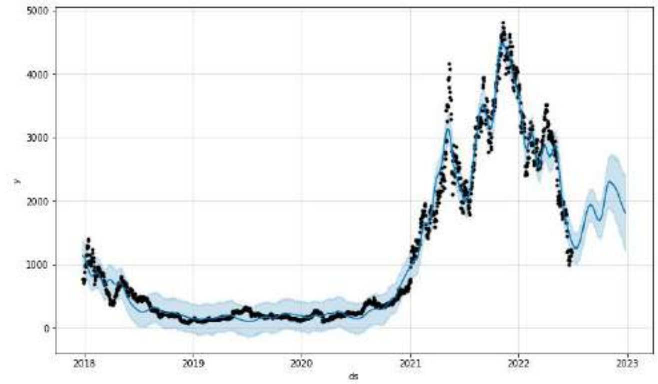


Figure 26: The figure of Prophet Model of Ethereum price result (9 – 1)

C. Litecoin

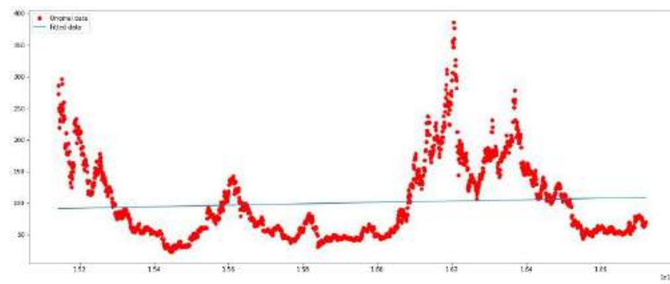


Figure 27: The figure of Linear Model of Litecoin price result

Figure 28: The figure of NonLinear Model of Litecoin price result

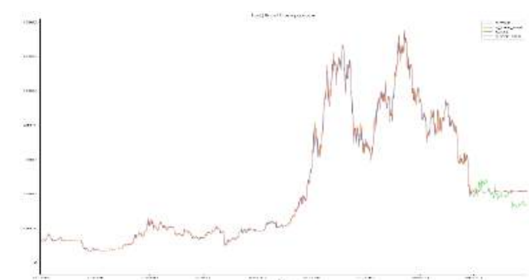


Figure 29: The figure of ARIMA Model of Litecoin price result (9 – 1)

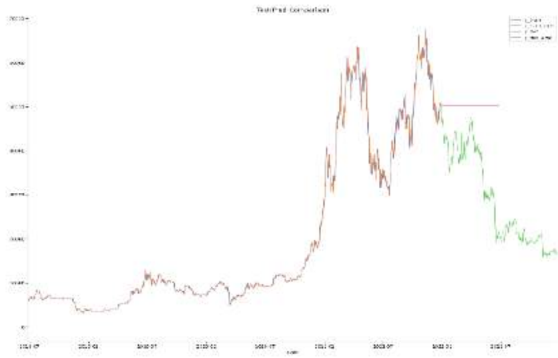


Figure 30: The figure of ARIMA Model of Litecoin price result (8 – 2)

Figure 31: The figure of ARIMA Model of Litecoin price result (7 – 3)

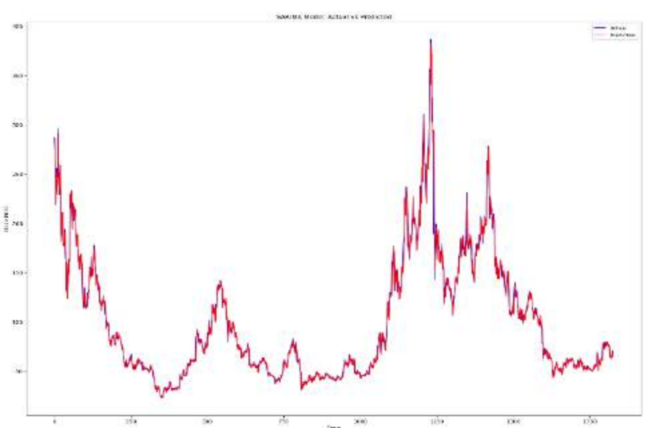


Figure 32: The figure of SARIMA Model of Litecoin price result (9 – 1)

Figure 33: The figure of SARIMA Model of Litecoin price result (8 – 2)

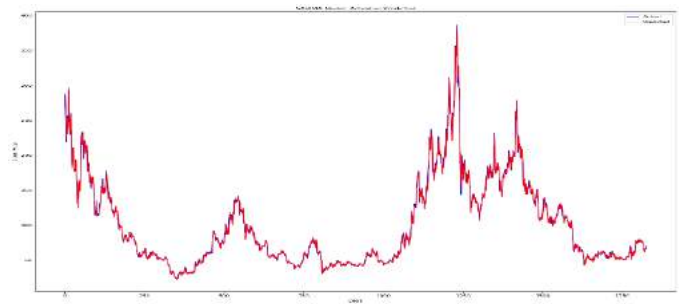


Figure 34: The figure of SARIMA Model of Litecoin price result (7 – 3)

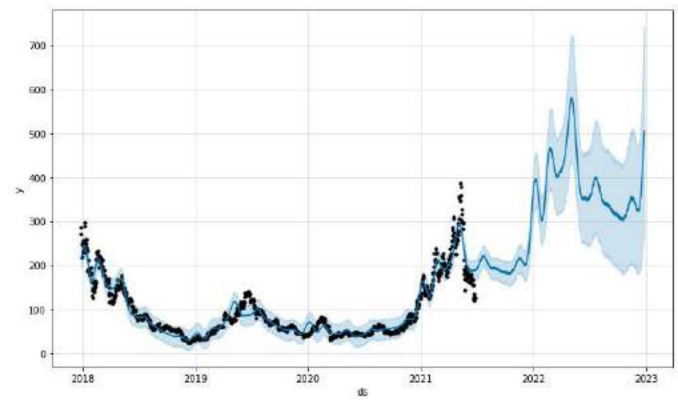


Figure 35: The figure of Prophet Model of Litecoin price result (7 – 3)

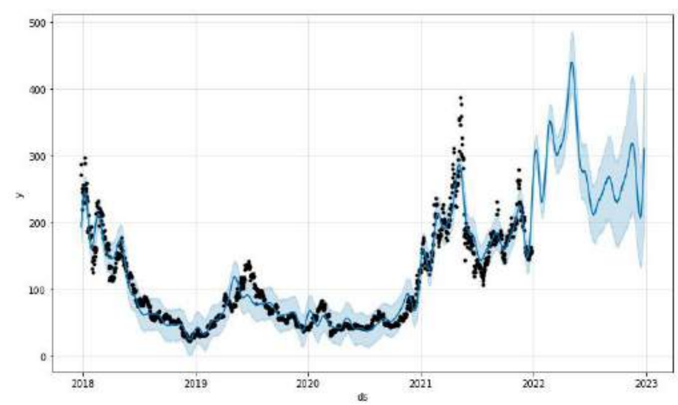


Figure 36: The figure of Prophet Model of Litecoin price result (8 – 2)

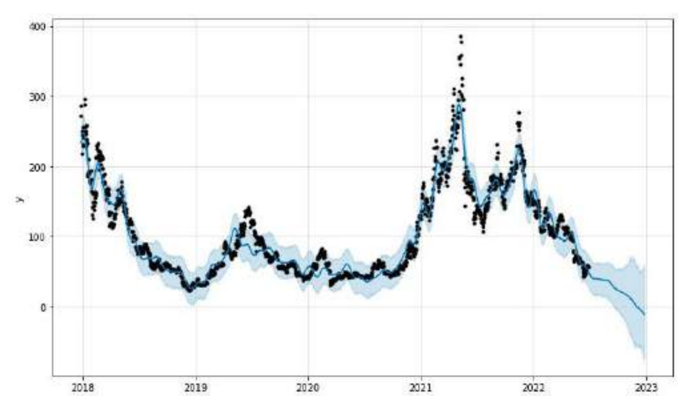


Figure 37: The figure of Prophet Model of Ethereum price result (9 – 1)

VII. CONCLUSION

In this study, we divided the training and testing data sets by 90% - 10%, 80% - 20% and 90% - 10% and statistics model and machine learning to predict the close price of three cryptocurrency. The ARIMA model learned more from training data, two linear learning such as LR and NLR has high errors. Besides, ARIMA model has learns good, but ARIMA still regression like LR, so the the predicted values has show a straight line. From that, it is difficult for predicting. And the excellent learning and small error prediction is LSTM model (Long Short Term Memory). Finally, with this study, we hope help the investor in invest the cryptocurrency.

REFERENCES

[1] Nicola Uras, Lodovica Marchesi, Michele Marchesi, Roberto Tonelli, Forecasting Bitcoin closing price series using linear regression and neural networks models, PeerJ Computer Science, 2020.

[2] Saad Ali Alahmari, USING NONLINEAR MACHINE LEARNING ALGORITHMS TO PREDICT THE PRICE OF CRYPTOCURRENCIES, International Journal of Future Generation Communication and Networking Vol. 13, No. 1, (2020).

[3] Poongodi M, Vijayakumar V, Naveen Chilamkurti, Bitcoin Price Prediction using ARIMA Model, International Journal of Web Based Communities, 2020.

[4] Xujun Zhang PhD, Yuanyuan Pang MD , Mengjing Cui MD , Lorann Stallones MPH, PhD , Huiyun Xiang MD, MPH, PhD, Forecasting mortality of road traffic injuries in China using seasonal autoregressive integrated moving average model, Annals of epidemiology, 2015

[5] Xin Huang , Wenbin Zhang, Xuejiao Tang, Mingli Zhang , Jayachander Surbiryala , Vasileios Iosifidis, Zhen Liu and Ji Zhang, LSTM Based Sentiment Analysis for Cryptocurrency Prediction, Journal of Soft Computing Paradigm 3.3, 2021.

[6] Is¸ıl Yenidogan, Aykut C¸ ayır ˘, Ozan Kozan, Tugc¸e Da ˘ g, C¸ i ˘ gdem Arslan, Bitcoin Forecasting Using ARIMA and PROPHET, 2018 3rd International Conference on Computer Science and Engineering (UBMK). IEEE, 2018.

[7] Reaz Chowdhury, M. Arifur Rahman, M. Sohel Rahman, M.R.C. Mahdy Predicting and Forecasting the Price of Constituents and Index of Cryptocurrency Using Machine Learning, Physica A: Statistical Mechanics and its Applications, 2020.

[8] Sean McNally, Jason Roche, Simon Caton, Predicting the Price of Bitcoin Using Machine Learning, 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, 2018.

[9] V. Derbentseva, V. Babenko, K. Khrustalevc, H. Obruchd, S. Khrustalovac, Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices, International Journal of Engineering, 2021.

[10] Azeez A.Oyedele, Anuoluwapo O.Ajayi, Lukumon O.Oyedele, Sururah A.Bello, Kudirat O.Jimoh, Performance evaluation of deep learning and boosted trees for cryptocurrency closing price prediction, Expert Systems With Applications, 2023.

[11] Mohammed Mudassir, Shada Bennbaia, Devrim Unal & Mohammad Hammoudeh, Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach, Neural Computing and Applications, 2020.

[12] Firat Akba, , Ihsan Tolga Medeni, Mehmet Serdar Guzel, Iman Askerzade, Manipulator Detection in Cryptocurrency Markets Based on Forecasting Anomalies, IEEE Access, 2021.

[13] Negar Malekia, Alireza Nikoubina, Masoud Rabbania, Yasser Zeinalib, Bitcoin Price Prediction Based on Other Cryptocurrencies Using Machine Learning and Time Series Analysis, Scientia Iranica, 2020.

[14] Yan Li, Wei Dai, Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model, The journal of engineering, 2020.

[15] Ho, Alvin, R. Vatambeti, and S. K. , Bitcoin Price Prediction Using Machine Learning and Artificial Neural Network Model, Ravichandran, Indian Journal of Science and Technology, 2021.

[16] R. K. Jana, Indranil Ghosh, Debojyoti Das, A differential evolution-based regression framework for forecasting Bitcoin price, Annals of Operations Research, 2021.

[17] Mohammad Ali, Swakkhar Shatabda, A Data Selection Methodology to Train Linear Regression Model to Predict Bitcoin Price, 2nd ICAICT, 28-29 November 2020.

[18] Mehdi Khashei, Mehdi Bijari, Gholam AliRaissi Ardali, Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs), Neurocomputing 72.4-6, 2009.

[19] Carl Dinshaw, Reetu Jain, Syed Abou Iltaf Hussain. "Statistical Scrutiny of the Prediction Capability of Different Time Series Machine Learning Models in Forecasting Bitcoin Prices", OMOTEC, last accessed [30/12/2022].

[20] Hari Krishnan Andi, An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model, Journal of Soft Computing Paradigm (JSCP), 2021.

[21] Sean J Taylor and Benjamin Letham. Forecasting at scale. The American Statistician, 2017.