КРИПТОВАЛЮТЫН ХАНШИЙГ ARIMA БОЛОН SARIMAX ЗАГВАРААР ТААМАГЛАХ НЬ

Г.Лхамдулам*, С.Цолмон**

Хураангуй: Энэхүү судалгааны зорилго нь хугацаан цуваан шинжилгээнд хамгийн өргөн ашигладаг ARIMA болон SARIMAX загварт тулгуурлан зарим криптовалютын ханшийг урьдчилан таамаглах явдал юм. Эдгээр загваруудын параметрийг тодорхойлохдоо автокорреляцийн функц (ACF) болон хэсэгчилсэн автокорреляцийн функц (PACF) шинжилгээг хийсэн. Судалгаандаа бид Finance.yahoo.com болон CoinMarketCup.com веб сайтаас 2015 оны 9 сарын 9-өөс 2022 оны 10 сарын 31-ны өдрийг хүртэлх 2600 өдрийн мэдээлэлд тулгуурлан, python 3.10 программ дээр боловсруулалт хийсэн. Хугацааны хоцролтын оновчтой зэргийг тогтоосны дараа биткойны ханшийг таамаглах хамгийн оновчтой загвар нь ARIMA (0,1,0)(0,0,0)[0], харин эфириумийн ханшийн хувьд ARIMA(2,1,1)(0,0,0)[0] загвар нь хамгийн бага алдаатай гарсан байна.

Түлхүүр үгс: машин сургалт, хугацаан цуваа, криптограф, биткойн, эфириум

FORECASTING CRYPTOCURRENCY RATES USING ARIMA AND SARIMAX MODELS

G.Lkhamdulam*, S.Tsolmon**

Abstract: The objective of this study is to forecast the exchange rates of select cryptocurrencies using the widely recognized ARIMA and SARIMAX models for time series analysis. Autocorrelation function (ACF) and partial autocorrelation function (PACF) analyses were conducted to ascertain the parameters for these models. Our research is grounded in a dataset spanning 2600 days sourced from Finance.yahoo.com and the CoinMarketCap.com website, covering the period from September 9, 2015, to October 31, 2022. The data was processed using Python 3.10. Upon identifying the optimal time lag, the preferred model for predicting Bitcoin prices is ARIMA (0,1,0)(0,0,0)[0], while for Ethereum, the optimal model is ARIMA(2,1,1)(0,0,0)[0], demonstrating the lowest error.

Keywords: machine learning, time series, cryptography, bitcoin, ethereum

* School of Engineering and Economics, Mandakh University, (E-mail): lkhamdulam@mandakh.edu.mn
** School of Engineering and Economics, Mandakh University, (E-mail): tsolmon@mandakh.edu.mn

Introduction

Cryptocurrencies represent virtual and electronic forms of currency, constituting a subset of virtual currencies secured through cryptography (Ivanchenko, 2021). Typically, cryptocurrencies employ intricate cryptographic algorithms, necessitating a network of interconnected computers to execute complex mathematical operations (Chowdhury, 2019). Notably, these currencies operate in an unregulated environment, rendering them highly volatile (Ivanchenko, 2021).

The allure of this market lies in the fact that the technology employed in cryptocurrency mining offers a viable alternative to conventional markets, such as gold. Additionally, cryptocurrencies stand out from traditional currencies as there is incomplete information available about cash transactions and the total currency in circulation. While forecasting cryptocurrency price movements is challenging, it is not deemed impossible (Valencia, 2019).

In recent years, alongside the utilization of certain cryptocurrencies as official currencies and for international payments, there has been a surge in investor interest in capitalizing on exchange rate differentials. Consequently, a considerable amount of research has been conducted in the realm of predicting cryptocurrency prices (Ivanchenko, 2021).

The study will further investigate the precision of predicting Bitcoin prices. One-variable dynamic models, specifically ARIMA, are employed for time series modeling, utilizing this approach to assess and forecast exchange rates. Through this model, we can discern which algorithm exhibits a lesser deviation from the actual value, enhancing the accuracy of predicting and forecasting short-term exchange rate prices.

Cryptocurrency and Blockchain technology

Cryptocurrency is a digital currency created for transactional purposes across computer networks. It operates as a decentralized and independent form of currency, free from the oversight and control of central entities such as governments or banks. The reluctance of central authorities to internationally recognize cryptocurrencies is rooted in the concern that national currencies might experience a short-term weakening against these digital counterparts. (niss.gov.mn, 2018).

Cryptocurrencies lack a physical form, unlike traditional paper money, and are typically not governed by a central authority. In contrast to Central Bank Digital Currencies (CBDCs), cryptocurrencies commonly operate with decentralized control. A cryptocurrency is typically deemed centralized if it is generated or created prior to issuance, or if it is issued by a singular entity. In the case of decentralized control, each cryptocurrency functions on a blockchain or distributed ledger technology, serving as a public database for financial transactions (Reaz Chowdhury, M. Arifur Rahman , M. Sohel Rahman ,M.R.C. Mahdy, 2015).

The inaugural cryptocurrency, Bitcoin, was introduced as open-source software in 2009. By March 2022, the market boasted over 9,000 cryptocurrencies, with more than 70 of them commanding a market capitalization surpassing $1 billion. Various parameters of the four types of openly traded cryptocurrencies, including market supply volume, 24-hour transactions, new block creation time, and exchange rates, exhibit distinct trading patterns in the market. The new block creation time, reward amount, and algorithmic change procedures differ for each cryptocurrency. Additionally, the speed of hash checks is determined by the computing power of miners in the network. Detailed explanations for each of these currencies are provided below.

Moreover, the blockchain functions as a mechanism employing cryptography and encryption, utilizing a specialized mathematical algorithm for the creation and verification of a progressively expanding data structure. Essentially, new information is added without deleting previous data, forming a sequential chain of "transaction blocks" (gratanet.com, 2020).

Blockchain 1.0 is expected to serve as a fundamental application for money transfers, remittances, and digital payment systems, akin to conventional cash or cryptocurrencies. Furthermore, Blockchain 2.0 extends beyond simple money transactions, evolving into a platform capable of managing a wider array of financial derivatives, including stocks, bonds, mortgages, futures, smart assets, and certificates.

The adoption of Blockchain 3.0 represents a rapid and cost-effective integration of blockchain technology, facilitating competition with centralized financial institutions. Its impact extends beyond the financial sector to enhance data security in the health sector, utilizing distributed technology for improved data storage and transmission. In the logistics industry, it enhances control over goods and products, while in the electoral system, it introduces transparency. This adoption aims to increase accessibility while simultaneously fortifying security (learn.bybit.com).

Being a cryptocurrency, it operates without regulation from governments or organizations, and its information is encrypted with a widely distributed private key within the network. While it shares the foundation of blockchain technology with other cryptocurrencies, its programming sets it apart.

Research design and methodology

The data utilized in this study were gathered from Finance.yahoo.com and CoinMarketCap.com, covering daily data from September 9, 2015, to October 31, 2022, for the peak market values of Bitcoin (BTC) and Ethereum (ETH), two prominent cryptocurrencies. In forecasting the exchange rates for these cryptocurrencies, we employed ARIMA (Autoregressive Integrated Moving Average) and SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors) models, commonly used in time series analysis, through the Python 3.10 programming language.

The ARIMA model was initially introduced in Peter Whittle's thesis, "Hypothesis Testing in Time Series Analysis." It is widely regarded as the most popular method for financial forecasting. Its recognition further expanded in 1971 when it was prominently featured in a book by George E.P. Box and Gwilym Jenkins.

ARIMA models are segmented into Autoregressive (AR) and Moving Average (MA) components. The Autoregressive (AR) model operates on the principle of regressing the target variable on its past values, with the AR model equation taking a lagged form.

Y is expressed as a linear function of the preceding n values, wherein the value of n can be replaced by coefficients B0 and B1. These beta values are determined during the model fitting process. The resulting equation can be employed to forecast future values by making appropriate adjustments to the equation.

$$Y = B0 + B1 * Y\_\text{lag } 1 + B2^{*} Y\_\text{lag } 2 + \cdots + Bn * Y\_\text{lagn} \quad [1]$$

$$\begin{aligned} Y\_\text{forward1} = B0 + B1 * Y + B2^{*} Y - \text{lag } 1 \\ + \cdots Bn^{*} Y\_\text{lag } (n-1) \end{aligned} \quad [2]$$

A component of ARIMA involves the process of data stationarity, which is applied when the time series data exhibits non-stationarity, as depicted in the equation below. [3] This process assumes that future values of Y are linear functions of its past changes and that the values of Y must exhibit a constant mean and variance.

$$\begin{aligned} Y\_\text{forward } 1 - Y = B0 + B1 * (Y - Y\_\text{ lag } 1) \\ + B2^{*}(Y\_\text{lag } 1 - Y\_\text{lag } 2) + \cdots \end{aligned} \quad [3]$$

ARIMA(p,d,q) serves as the standard notation for representing ARIMA models. These parameters can be substituted with integers to define the specific model type. “p” signifies the number of lags of Y included in the model, “d” represents the order of differencing needed to achieve data stationarity, and “q” denotes the MA order, which is the number of backward prediction errors (Chahat, Tandon; Sanjana, Revankar; Hemant, Palivel, 2021).

Research results

The selected cryptocurrencies, Bitcoin (BTC) and Ethereum (ETH), are employed in forecasting cryptocurrency exchange rates through a sequential ARIMA model. This approach assumes that the daily exchange rate is dependent on its preceding value and a random error. Logarithms were applied to exhibit linearity in the analysis.

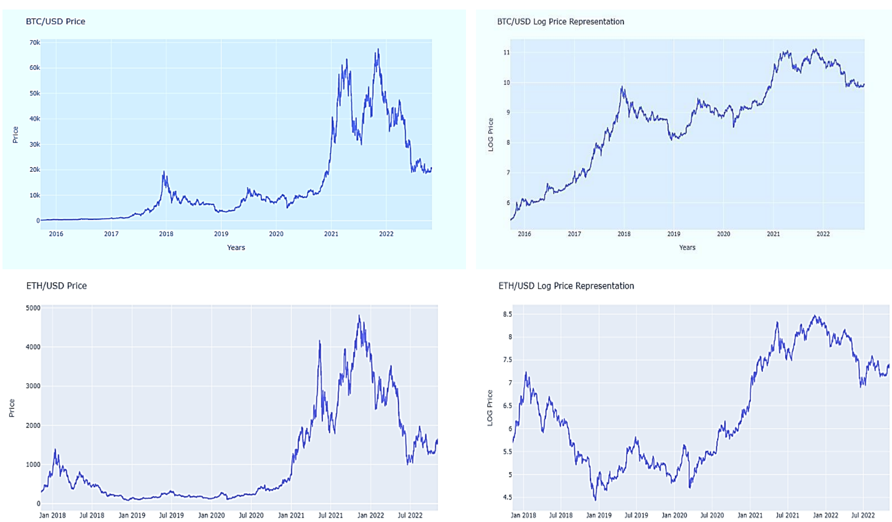

Figure 1 illustrates the price movements of Bitcoin and Ethereum.

Assessing the stability of the series involves examining whether the autocorrelation function has manifested in the time series, forming the basis for the reliability of the evaluation. Consequently, the stability of the exchange rate series for selected cryptocurrencies such as BTC and ETH is examined and presented in Figure 2.

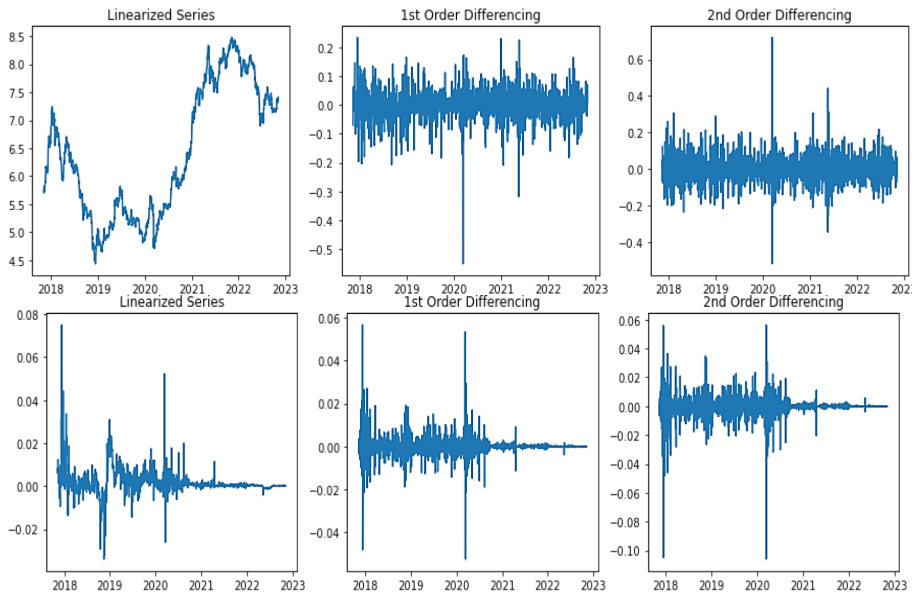

Figure 2 displays the outcomes of the rate stability testing.

To evaluate the series stability, two hypotheses were formulated: the null hypothesis (H0) positing "There is no instability in the series," and the alternative hypothesis asserting "There is instability in the series." These hypotheses were tested using the Augmented Dickey-Fuller test (ADF).

Table 1 presents the results of the Augmented Dickey-Fuller (ADF) test

Time series			t-Statistic	ρ-Value
		1%	-3.432868573	0.21
	BTC	5%	-2.862652711	0
		10%	-2.567362336	2.82E-29
		1%	-3.433948103	0.803137136
test critical value	ETH	5%	-2.863129362	0
		10%	-2.567616123	1.04E-26
		1%	-3.433974025	6.56E-06
	USDT	5%	-2.863140805	9.50E-28
		10%	-2.567622216	4.15E-29

Source: Researchers' estimates

Table 1 displays the augmented Dickey-Fuller (ADF) test t-statistics and corresponding p-values for cryptocurrency time series, along with transformed time series, at 1%, 5%, and 10% significance levels. Both BTC and ETH time series exhibit a unit root, indicating that these series are not constant.

The p-value (<0.05) for BTC provides strong evidence against the H0 hypothesis at the 1-5% significance levels for both the linear series and the 1st-order differenced series. Therefore, the H0 hypothesis is rejected (indicating consistency). However, the 2nd-order differenced series has a p-value (>0.05) at the 10% significance level, suggesting weak evidence against the H0 hypothesis, and thus, the H0 hypothesis cannot be rejected (indicating instability). For ETH, the p-value (<0.05) provides strong evidence against the H0 hypothesis at the 5% significance level, as the 1st-order differenced series leads to the rejection of the H0 hypothesis (consistent). In the case of both BTC and ETH cryptocurrencies, the optimal differencing parameter (using n diffs) was determined using pmdarima, and the 1st-order difference was identified as the best-performing.

Cryptocurrencies BTC and ETH are elucidated using the ARIMA model within the framework of a time series model. The accompanying graph illustrates that the Autocorrelation Function (ACF) values for both BTC and ETH fall within the 95% confidence interval (depicted by the dotted gray line or remain below the gray line) for lags greater than 0. This observation affirms the absence of autocorrelation in our data. Typically, exchange rates from one day are highly dependent on those of the previous day, implying a dependence on their predecessors. Crossing the blue-gray line of the confidence interval indicates statistical significance or autocorrelation.

Table 1. SARIMA model results for BTC

<table>
<tr><td>Dep.Variable:</td><td colspan="2">Close</td><td colspan="2">No.Observations:</td><td colspan="2">2600</td></tr>
<tr><td>Model:</td><td colspan="2">ARIMA(1, 1, 1)</td><td colspan="2">Log Likelihood</td><td colspan="2">4737.254</td></tr>
<tr><td>Date:</td><td colspan="2">Tue, 01 Nov 2022</td><td colspan="2">AIC</td><td colspan="2">-9468.509</td></tr>
<tr><td>Time:</td><td colspan="2">7:54:57</td><td colspan="2">BIC</td><td colspan="2">-9450.92</td></tr>
<tr><td>Sample:</td><td colspan="2">9/19/2015</td><td colspan="2">HQIC</td><td colspan="2">-9462.136</td></tr>
<tr><td rowspan="2">Covariance Type:</td><td colspan="6">org</td></tr>
<tr><td>coef</td><td>std err</td><td>z</td><td>P>|z|</td><td>[0.025]</td><td>[0.975]</td></tr>
<tr><td>ar.L1</td><td>-0.3827</td><td>0.492</td><td>-0.778</td><td>0.437</td><td>-1.348</td><td>0.582</td></tr>
<tr><td>ma.L1</td><td>0.3562</td><td>0.496</td><td>0.718</td><td>0.473</td><td>-0.616</td><td>1.328</td></tr>
<tr><td>sigma2</td><td>0.0015</td><td>1.68E-05</td><td>90.954</td><td>0</td><td>0.001</td><td>0.002</td></tr>
</table>

Ljung-Box (L1) (Q):	0.04	Jarque-Bera (JB):	13451.71
Prob(Q):	0.84	Prob(JB):	0
Heteroskedasticity (H):	0.81	Skew:	-0.73
Prob(H) (two-sided):	0	Kurtosis:	14.05

Source: Researchers' estimates

The outcomes of the SARIMA model indicate that the coefficients of the Autoregressive (AR) and Moving Average (MA) terms are both less than 1 and statistically significant, with p-values below 0.05. This leads to the selection of the SARIMA model, specifically ARIMA(1,1,1) for BTC and ARIMA(1,1,1) for ETH, as the most suitable models for forecasting.

Table 2. SARIMAX model results for ETH

Dep.Variable:	Close		No.Observations:		1823	
Model:	ARIMA(1, 1, 1)		Log Likelihood		2823.988	
Date:	Sat, 05 Nov 2022		AIC		-5641.975	
Time:	12:17:40		BIC		-5625.452	
Sample:	11/9/2017		HQIC		-5635.88	
Covariance Type:	org					
	coef	std err	z	P>|z|	[0.025]	[0.975]
ar.L1	-0.7623	0.129	-5.905	0	-1.015	-0.509
ma.L1	0.7172	0.139	5.155	0	0.444	0.99
sigma2	0.0026	3.69E-05	71.412	0	0.003	0.003
Ljung-Box (L1) (Q):	0.07		Jarque-Bera (JB):		7293.11	
Prob(Q):	0.8		Prob(JB):		0	
Heteroskedasticity (H):	0.83		Skew:		-0.93	
Prob(H) (two-sided):	0.02		Kurtosis:		12.62	

Source: Researchers' estimates

Now that the BTC and ETH time series data are preprocessed, the next step involves determining the optimal p, d, q values for the ARIMA model. This is accomplished using the auto_arima function, which calculates the most suitable values of p, d, q to optimize the predictive performance of the ARIMA model. The optimal predictive model is identified based on the lowest Akaike Information Criterion (AIC) value. After employing the auto_arima function to determine optimal values, the best model for predicting the BTC rate is identified as ARIMA(0,1,0)(0,0,0)[0] with an intercept, while for the ETH rate, ARIMA(2,1,1)(0,0,0)[0] is identified as the optimal model for the dataset.

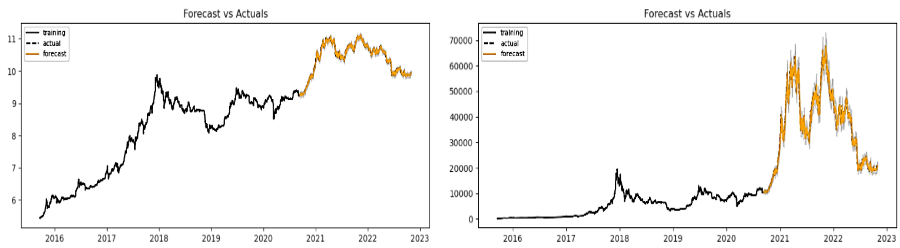

Figure 3 displays the actual values of BTC and the corresponding exchange rate predictions.

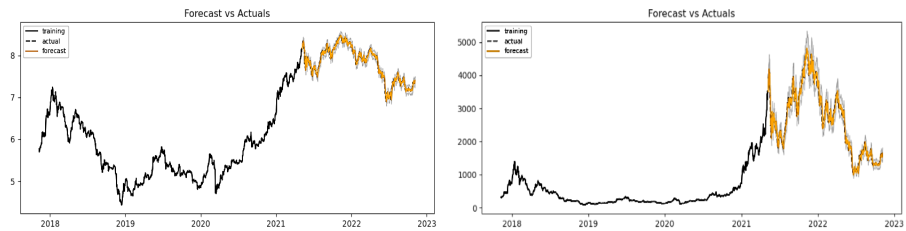

Figure 4 illustrates the real values of ETH alongside the corresponding rate predictions.

Following the determination of the optimal p, d, q values for the ARIMA model, the time series proceeds to the training and testing phase. The dataset is divided into a 70:30 ratio, with 70% used for training and the remaining 30% for testing the model. The model is executed, generating a model_prediction object for subsequent predictions. During this process, consider creating a visualization such as "Plot forecast vs. real," which includes the historical training set for better understanding and analysis.

Once the ARIMA model has been trained and future predictions have been tested, the np.exp(ARIMA.predicted_mean) function is employed. This function, with parameters including the starting value and length, is used to predict the next 5 days of BTC and ETH rates. The application of np.exp suggests that the predictions are based on the exponential transformation of the predicted mean.

```
In [92]: print(np.exp(ARIMA(log_df,order=best_order).fit().get_forecast(5).predicted_mean))

2022-11-07    21248.128906
2022-11-08    21248.128906
2022-11-09    21248.128906
2022-11-10    21248.128906
2022-11-11    21248.128906
Freq: D, Name: predicted_mean, dtype: float64
```

Figure 5 illustrates the 5-day prediction of BTC closing rates.

```
In [40]: print(np.exp(ARIMA(log_df,order=best_order).fit().get_forecast(5).predicted_mean))

2022-11-07    1631.481823
2022-11-08    1633.056827
2022-11-09    1634.172578
2022-11-10    1634.935799
2022-11-11    1635.461046
Freq: D, Name: predicted_mean, dtype: float64
```

Figure 6 illustrates the 5-day prediction of ETH closing rates.

The final step in the forecasting process involves evaluating the model's performance. This assessment can be conducted using various estimation methods, such as mean square error (MSE), mean absolute error (MAE), and root mean square error (RMSE). Figures 7 and 8 present the assessments made using mean absolute percentage error (MAPE) and mean absolute scaled error (MASE). The evaluation includes:

```
In [90]: # MAPE and MASE
         print('mape model:', get_mape(result.Close, result.forecast))
         print('mape baseline:', get_mape(result.Close, result.base))
         print('')
         print('mase model:', get_mase(result.Close, result.forecast, y_train.Close))
         print('mase baseline', get_mase(result.Close, result.base, y_train.Close))

mape model: 0.26
mape baseline: 0.26

mase model: 1.06
mase baseline 1.06
```

Figure 7 displays the results of the evaluation for BTC using Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE).

```
In [38]: # MAPE and MASE
         print('mape model:', get_mape(result.Close, result.forecast))
         print('mape baseline:', get_mape(result.Close, result.base))
         print('')
         print('mase model:', get_mase(result.Close, result.forecast, y_train.Close))
         print('mase baseline', get_mase(result.Close, result.base, y_train.Close))

mape model: 0.47
mape baseline: 0.46

mase model: 1.02
mase baseline 1.01
```

Figure 8 displays the results of the evaluation for ETH using Mean Absolute Percentage Error (MAPE) and Mean Absolute Scaled Error (MASE).

Conclusion

In this research, the objective is to forecast the prices of major market players, BTC and ETH, utilizing ARIMA and SARIMAX models. Initially, the data was examined and stabilized to ensure its stability. According to the study results, 87 percent of short-term exchange rate movements were accurately predicted. Furthermore, the evaluation of BTC and ETH cryptocurrencies involves assessing mean absolute percentage error (MAPE) and mean absolute scaled error (MASE). The irregular patterns in Bitcoin's price movements present challenges in determining an optimal ARIMA model. A model effective for forecasting in one time range may not necessarily perform well in the subsequent time range. There are plans to incorporate more parameters and explore additional time series models to enhance the accuracy of predicting Bitcoin prices.

References

Chahat, Tandon; Sanjana, Revankar; Hemant, Palivel. (2021). How can we predict the impact of the social media messages on the value.

Chowdhury, R. (2019). Predicting and Forecasting the Price of Constituents and Index of.

Ivanchenko, I. S. (2021). Fundamentals of forecasting cryptocurrence rates.

Reaz Chowdhury, M. Arifur Rahman , M. Sohel Rahman ,M.R.C. Mahdy. (2015). Predicting and Forecasting the Price of Constituents and Index of.

Valencia, F. (2019). Price Movement Prediction of Cryptocurrencies UsingSentiment Analysis and Machine Learning.

Xianghua Wu, Jieqin Zhou , Huaying Yu. (2021). The Development of a Hybrid Wavelet-ARIMA-LSTM Model for Precipitation Amounts and Drought Analysis.