Prediction of Bitcoin Pricing by SHapley Additive exPlanations

Danting Tao, Yitong Wu, Yutong Liu*

Wenzhou-Kean University, Wenzhou, Zhejiang, China

*Corresponding author: sigmund307@gmail.com

Keywords: Bitcoin, Prediction Model, SHapley Additive exPlanations

Abstract: This paper investigates the model that predicts the bitcoin price trend and uses the SHapley Additive exPlanations method to explain the model predictions. The article first introduces the background and development of the cryptocurrency market, and then outlines the relevant literature review. In the experimental part, we use various features and data to train the random forest classifier and interpret its predictions using SHAP values. Finally, it is concluded that BTC_Momentum is one of the most critical factors affecting the bitcoin price trend.

1. Introduction

About a decade ago, a brand-new market called cryptocurrency was established. The establishment has something to do with the global financial crisis. Cryptocurrency is a by-product of blockchain. Bitcoin is developed based on blockchain. Cryptocurrency is a medium of exchange that ensures trade's safety. Nowadays, there are three kinds of most famous cryptocurrencies in the world. They are Bitcoin, Ether, and Litecoin. Among the three, Bitcoin is the most renowned cryptocurrency. Bitcoin is very different from traditional currencies. At present, many countries in the world do not yet regard Bitcoin as a legal means of payment. Bitcoin is a virtual digital currency. The cryptocurrency market is decentralized, which means it can allow people anywhere to trust each other and carry out transactions via the Internet without a central management institution. Cryptocurrency is virtual, not real, and it is carried out on a platform. Our research is about the trend of Bitcoin's price. Because nowadays, cryptocurrency is becoming more and more important. At the same time, the cryptocurrency market faces the risk of high volatility as well as pricing bubbles. Our state-of-the-art forecasts of the price trend of Bitcoin have important implications for a diverse range of stakeholders, including investors, financial institutions, policymakers, and academia. For investors, forecasting the price trend of Bitcoin can help investors earn more profits. Investors can buy Bitcoin when the price rises, and then sell it when the price falls, in order to gain more profits. By predicting the trend of Bitcoin prices, we can understand the market's views on the development of digital currencies. When investors are confident in digital currencies, they will buy more Bitcoin to drive up prices. When investors are skeptical about digital currencies, they may reduce their investment in Bitcoin.

2. Literature review

Nowadays, there are so many papers about the cryptocurrency. The research fields vary from single concepts to relevant psychology investments. We chose ten pieces of pertinent research as our literature review.

2.1 Cryptocurrency Mechanisms for Blockchains: Models, Characteristics, Challenges, and Applications [1]

This research briefly introduces the relationship between cryptocurrency and blockchain, some related frameworks, and other essential information as its title.

  DOI: 10.25236/iceesr.2024.047

2.2 Financial literacy or investment experience: which is more influential in cryptocurrency investment? [2]

This study is devoted to exploring the determined factor in the investment field of cryptocurrency. And they finally find that investment experience has greater power than financial literacy for the final decision-making process.

2.3 The Challenge of Cryptocurrency in the Era of the Digital Revolution: A Review of Systematic Literature [3]

The research provides a clear explanation of some concepts in the cryptocurrency field and analyzes the working principle and threat of blockchain technology.

2.4 The psychology of cryptocurrency trading: Risk and protective factors [4]

This research doesn't follow traditional research topics. It uses psychological knowledge to break down the logical relationship between the market volatility of cryptocurrency and public opinion and holder mentality to demonstrate the influence individually.

2.5 Cryptocurrency Scams: Analysis and Perspectives [5]

Focused on analyzing the behavior of each cryptocurrency scam, this Italian team dissected the scam examples and classified them into different categories. The foremost is that it provides a brand-new thought to prevent fake information about cryptocurrency.

2.6 An Advanced CNN-LSTM Model for Cryptocurrency Forecasting [6]

This research introduced a new forecast model which can handle various cryptocurrency data individually. The deep neural network makes this new model more precious than former models.

2.7 Cryptocurrency trading: a comprehensive survey [7]

This research makes a generalized analysis of cryptocurrency from the transaction technology, transaction system, market condition, miscellaneous research, portfolio, and relevant assets.

2.8 Bitcoin and Cryptocurrency: Challenges, Opportunities and Future Works [8]

The corresponding authors survey the law and regulation, energy consumption, and relevant attacks to analyze the basic information that has been taken into consideration by some regular investors.

2.9 The Pricing and Performance of Cryptocurrency [9]

This research is devoted to acquiring ICO pricing and a series of long-term pricing of some new cryptocurrencies to determine the relationship between the pricing strategies and the potential variations of some new cryptocurrencies.

2.10 Cryptocurrency Blockchain Technology in the Digital Revolution Era [10]

This research interprets the blockchain technique from the currency and digital technology function.

3. Code description

3.1 Feature

3.1.1 Illiquidity measure and returns

At first, with the intention of calculating daily returns, we used the closing price (BTC_Closing) as the return rate (BTC_Return). In addition, we used the return rate (BTC_Return) to divide the relevant transaction volume (BTC_Volume) as the illiquidity data (BTC_Illiquidity). We began the research from 2018/2/1 to 2018/2/5, and got 4 groups of return rate and illiquidity data in table 1.

Table 1 Illiquidity measure and returns

	Date	Value	Value_Classification	BTC_Closing	BTC_Volume	EPU	GPR	BTC_Return	BTC_Illiquidity
0	2018/2/1	30.0	Fear	9170.540039	9.959400e+09	49.78	87.301842	NaN	NaN
1	2018/2/2	15.0	Extreme Fear	8830.750000	1.272690e+10	61.20	83.734215	-0.037052	2.911341e-12
2	2018/2/3	40.0	Fear	9174.910156	7.263790e+09	44.00	81.256027	0.038973	5.365370e-12
3	2018/2/4	24.0	Extreme Fear	8277.009766	7.073550e+09	145.10	80.819733	-0.097865	1.383531e-11
4	2018/2/5	11.0	Extreme Fear	6955.270020	9.285290e+09	145.79	81.007408	-0.159688	1.719796e-11

3.1.2 1-week Momentum

For the purpose of calculating the daily momentum (BTC_Momentum), we chose 1 week as a period and used the daily closing price minus the closing price of the day before. We began the research from 2018/2/1 to 2018/2/10, and have 9 daily momentum data. And we have 3 valid momentum data in table 2.

Table 2 1-week Momentum

	Date	Value	Value_Classification	BTC_Closing	BTC_Volume	EPU	GPR	BTC_Return	BTC_Illiquidity	BTC_Momentum
0	2018/2/1	30.0	Fear	9170.540039	9.959400e+09	49.78	87.301842	NaN	NaN	NaN
1	2018/2/2	15.0	Extreme Fear	8830.750000	1.272690e+10	61.20	83.734215	-0.037052	2.911341e-12	NaN
2	2018/2/3	40.0	Fear	9174.910156	7.263790e+09	44.00	81.256027	0.038973	5.365370e-12	NaN
3	2018/2/4	24.0	Extreme Fear	8277.009766	7.073550e+09	145.10	80.819733	-0.097865	1.383531e-11	NaN
4	2018/2/5	11.0	Extreme Fear	6955.270020	9.285290e+09	145.79	81.007408	-0.159688	1.719796e-11	NaN
5	2018/2/6	8.0	Extreme Fear	7754.000000	1.399980e+10	92.59	81.883293	0.114838	8.202838e-12	NaN
6	2018/2/7	36.0	Fear	7621.299805	9.169280e+09	58.29	81.692352	-0.017114	1.866425e-12	NaN
7	2018/2/8	30.0	Fear	8265.589844	9.346750e+09	89.29	82.237068	0.084538	9.044649e-12	-904.950195
8	2018/2/9	44.0	Fear	8736.980469	6.784820e+09	82.68	82.249718	0.057030	8.405601e-12	-93.769531
9	2018/2/10	54.0	Neutral	8621.900391	7.780960e+09	138.06	80.279343	-0.013172	1.692800e-12	-553.009765

3.1.3 Turnover Ratio

We used the daily transaction volume (BTC_Volume) to divide the daily closing price (BTC_Closing) to get the turnover ratio. We chose the data from 2018/2/1 to 2018/2/5 as the dataset. At last, all 5 data became valid data in table 3.

Table 3 Turnover Ratio

	Date	Value	Value_Classification	BTC_Closing	BTC_Volume	EPU	GPR	BTC_Return	BTC_Illiquidity	BTC_Momentum	BTC_Turnover_Ratio
0	2018/2/1	30.0	Fear	9170.540039	9.959400e+09	49.78	87.301842	NaN	NaN	NaN	1.086021e+06
1	2018/2/2	15.0	Extreme Fear	8830.750000	1.272690e+10	61.20	83.734215	-0.037052	2.911341e-12	NaN	1.441203e+06
2	2018/2/3	40.0	Fear	9174.910156	7.263790e+09	44.00	81.256027	0.038973	5.365370e-12	NaN	7.917015e+05
3	2018/2/4	24.0	Extreme Fear	8277.009766	7.073550e+09	145.10	80.819733	-0.097865	1.383531e-11	NaN	8.546021e+05
4	2018/2/5	11.0	Extreme Fear	6955.270020	9.285290e+09	145.79	81.007408	-0.159688	1.719796e-11	NaN	1.335001e+06

3.1.4 Moving Average

With the intention of decreasing sporadic uncertainty and finding the trend of the daily closing price (BTC_Closing), we chose the date from 2018/2/1 to 2018/2/10 and calculated the 7-day moving average of the daily closing price. Finally, we got 4 valid data seen in Table 4.

Table 4 Moving Average

	Date	BTC_Closing	7_day_MA
0	2018/2/1	9170.540039	NaN
1	2018/2/2	8830.750000	NaN
2	2018/2/3	9174.910156	NaN
3	2018/2/4	8277.009766	NaN
4	2018/2/5	6955.270020	NaN
5	2018/2/6	7754.000000	NaN
6	2018/2/7	7621.299805	8254.825684
7	2018/2/8	8265.589844	8125.547084
8	2018/2/9	8736.980469	8112.151437
9	2018/2/10	8621.900391	8033.150042

3.1.5 Exponential MA

Traditional moving average (MA) has apparent hysteresis, which means it has limited value of reference. We want to predict the variation trend smoothly and precisely, so we calculated the 7-day Exponential Moving Average (EMA). Finally, we got 10 valid data as shown in table 5.

Table 5 Exponential MA

	Date	BTC_Closing	7_day_MA	7_day_EMA
0	2018/2/1	9170.540039	NaN	9170.540039
1	2018/2/2	8830.750000	NaN	9085.592529
2	2018/2/3	9174.910156	NaN	9107.921936
3	2018/2/4	8277.009766	NaN	8900.193893
4	2018/2/5	6955.270020	NaN	8413.962925
5	2018/2/6	7754.000000	NaN	8248.972194
6	2018/2/7	7621.299805	8254.825684	8092.054097
7	2018/2/8	8265.589844	8125.547084	8135.438033
8	2018/2/9	8736.980469	8112.151437	8285.823642
9	2018/2/10	8621.900391	8033.150042	8369.842830

3.2 Exploratory Data Analysis

3.2.1 Visualization

We visualize Bitcoin Closing Prices Over Time from 2018 to 2023. We set the horizontal coordinate to the data of time, set the ordinate to the closing price of Bitcoin. Thus, the curve shows how the closing price of Bitcoin changes over time. And based on the chart, we can see that the price of Bitcoin was at a lower level around 2018, around $3,000. Then, starting in 2019, the price of Bitcoin began to gradually rise, reaching a peak in 2020, around $40,000. However, by 2021, the price of Bitcoin began to show significant volatility, falling rapidly from its peak to around $27,000 by the end of the year. Going into 2022, the price of Bitcoin continued to fall, and by the end of the year it had fallen to around $15,000. For the forecast of 2023, according to the chart, the price of Bitcoin could fluctuate between $15,000 and $25,000. (Figure 1)

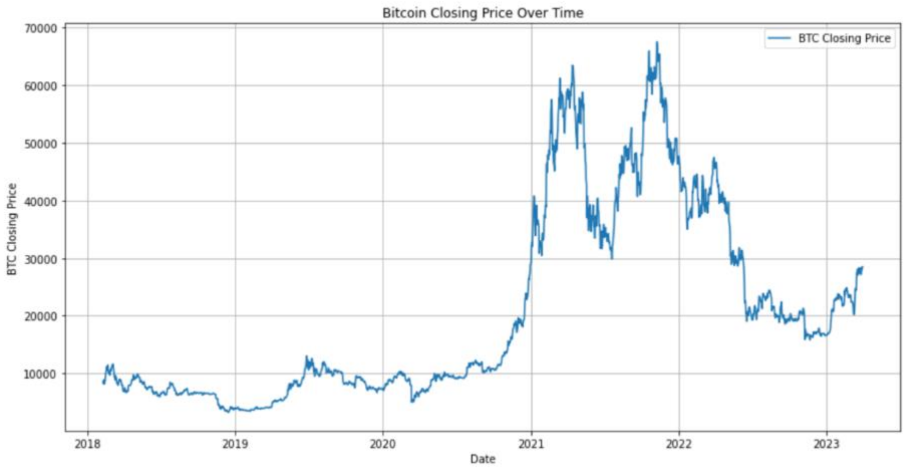

Figure 1 Bitcoin Closing Price Over Time

There are two types of price trend of Bitcoin. Due to micro market friction, we chose 0.0025 as the boundary for the rise and fall of Bitcoin. When BTC_Return ＞ 0.0025， we mark BTC_Return_Label as 1. When BTC_Return≤0.0025，we mark BTC_Return_Label as 0. Then we should calculate the proportion of 1 and 0 in BTC_Return_Label. Finally, we find the result is 47 to 53. Because our final proportion is 47 to 53, there is no need for non-imbalance testing. (Table 6)

Table 6 Binary variable for BTC_Return

	Date	BTC_Return	BTC_Return_Label
7	2018-02-08	0.084538	1
8	2018-02-09	0.057030	1
9	2018-02-10	-0.013172	0
10	2018-02-11	-0.057056	0
11	2018-02-12	0.097983	1

After that, we recalculate the correlation matrix and plot a heatmap for the correlation matrix. Due to the 7_day_MA data is approximately equal to the 7_day_EMA data and it's useless to have the two same parameters, we intend to give up the 7_day_EMA data. (Figure 2)

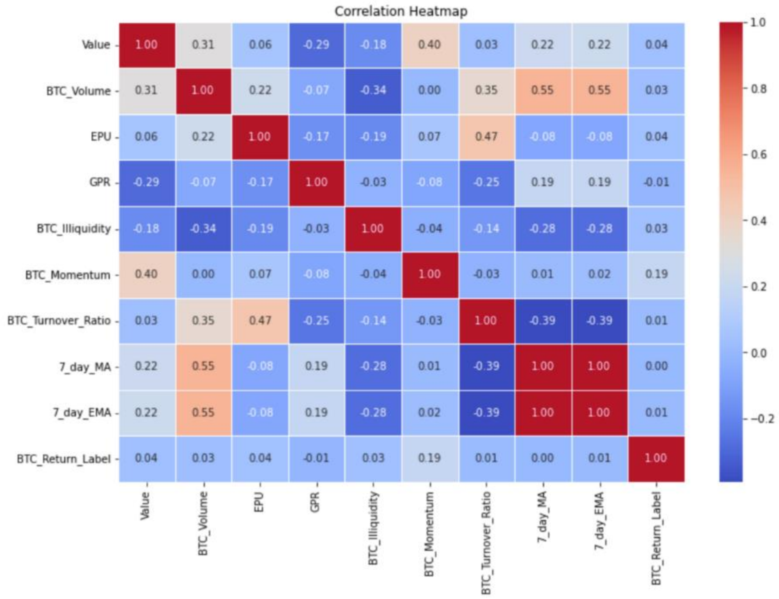

Figure 2 Correlation Heatmap

3.3 Random Forest Classifier

We imported the value, the BTC_Volume, the EPU, the GPR, the BTC_Illiquidity, the BTC_Momentum, the 7_day_MA, and the BTC_Turnover Ratio from formal research outcomes to generate the BTC_Return_Label. We set the test size equal to 0.2, and the random state equal to 42 to start the training process. Then we used the accuracy score, classification report and confusion matrix to make an evaluation. In addition, we introduced the ROC curve, confusion matrix and classification report. By the calculation, we find that the ROC curve area is equal to 0.62, which means this research method has a positive meaning. In our confusion matrix, our predicted label is closer to the true label, which means our data is useful and valid. (Figure 3, Figure 4)

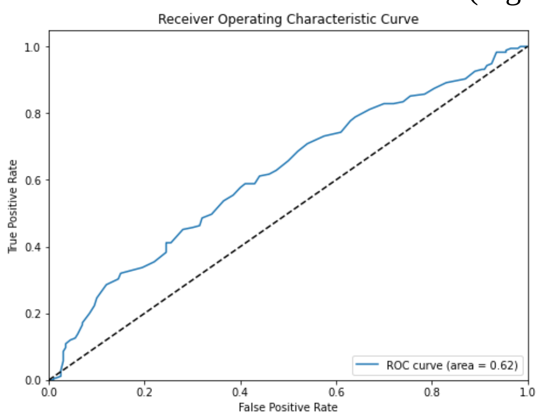

Figure 3 Receiver Operating Characteristic Curve

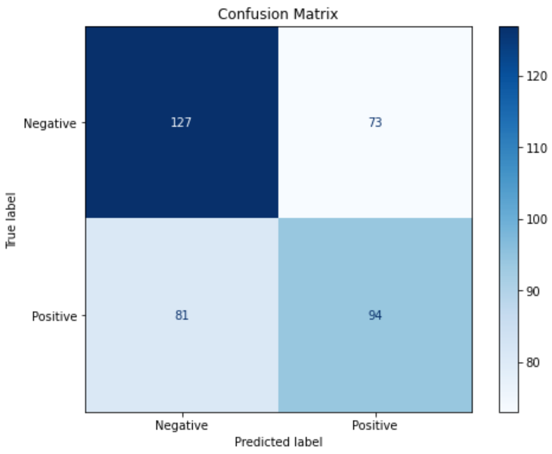

Figure 4 Confusion Matrix

3.3.1 Grid Search

We used the GridSearchCV to Exhaust the relevant parameter. Finally, we got the best parameters from the grid search. At last, the max depth is 10, the max feature is auto, the min samples leaf is 1, the min samples split is 10, and the n estimator is 200. Our best score is approximately equal to 0.5973.

3.4 Retrain the model with optimal parameters

The code Retrain the model with optimal parameters defines a training predictor function and takes three parameters: the training data, the test data, and the target variable. Inside the function, the training data is first split, with 80% of the data used for training and 20% for cross-validation. Then, the random forest classifier is trained using training data and cross-validation data. Next, the trained classifier is used to make predictions on the test data and the predictions are saved. Finally, the function returns the real and predicted labels for the test data. And we get the result BTC_Momentum is the most important parameters. Then, Predict the probabilities of positive class, compute the ROC-AUC and compute the ROC curve. We figure the ROC curve and named it Receiver Operating Characteristic Curve, and set the X coordinate to False Positive Rate and set the Y coordinate to True Positive Rate. Finally, we set the legend, save and show the image. The curve from left to right indicates that the false positive rate increases gradually while the true positive rate remains unchanged. Or the true positive rate gradually increases and the false positive rate remains the same. By looking at the ROC curve, we can evaluate the performance of the classifier. Ideally, the ROC curve will converge to the upper left corner, indicating a high true positive rate and a low false positive rate. The closer the curve is to the top left corner, the better the classifier performs. We can draw the following conclusion: the classifier has a strong ability to identify positive samples, but there may be cases where negative samples are misjudged as positive samples. (Figure 5)

In the resulting confusion matrix, we can see that the blue part represents the number of negative classes that the model incorrectly predicted when the actual label was positive; The darker part represents the number of positive classes that the model incorrectly predicted when the actual label was negative. By comparing the sizes of these two parts, we can assess the accuracy of the model in distinguishing between positive and negative classes. (Figure 6)

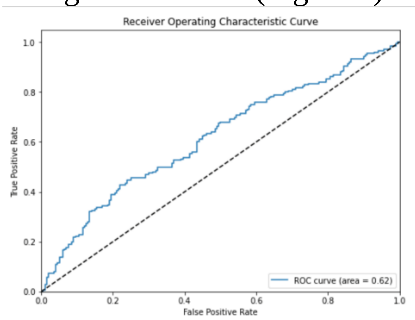

Figure 5 Receiver Operating Characteristic Curve 2.0

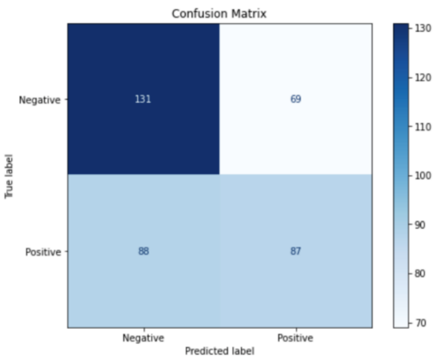

Figure 6 Confusion Matrix 2.0

4. Results-SHapley Additive exPlanations

In the end, we explain the model's predictions using SHapley values.

Higher stands for it have positive effect on the model, lower stands for it has negative effect on the model. We preliminary predict that the volatility trend of bitcoin has the highest correlation with the BTC_Momentum, as high as 0.91. (Figure 7)

Figure 7 Preliminary prediction

After optimizing the model, the accuracy of the model decreases, but we still get the volatility trend of bitcoin has the highest correlation with the BTC_Momentum, as high as 0.77. (Figure 8)

Figure 8 Optimized prediction

5. Conclusion

We visualize Bitcoin Closing Prices Over Time from 2018 to 2023. The curve shows how the closing price of Bitcoin changes over time. Then we use 0 and 1 to mark the trend of Bitcoin, and we choose 0.0025 as the boundary for the rise and fall of Bitcoin. The final result is 47 to 53. We calculate daily returns, the daily momentum (BTC_Momentum) and get the turnover ratio. In addition, we introduced the ROC curve, confusion matrix and classification report. By the calculation, we find that the ROC curve area is equal to 0.62. We used the GridSearchCV to Exhaust the relevant parameter. Finally, we got the best parameters from the grid search. The code Retrain the model with optimal parameters defines a training predictor function and takes three parameters: the training data, the test data, and the target variable. We can draw the following conclusion: the classifier has a strong ability to identify positive samples, but there may be cases where negative samples are misjudged as positive samples. In the resulting confusion matrix, by comparing the sizes of these two parts, we can assess the accuracy of the model in distinguishing between positive and negative classes. In the end, explain the model's predictions using SHapley values.

In addition to studying the price trend of Bitcoin, we may also study the supply of Bitcoin in the future. The supply of Bitcoin is limited, so in the future we may study its value storage function. As the first successful cryptocurrency, Bitcoin still has a lot of research space in the future.

References

[1] Sharma, D. K., Pant, S., Sharma, M., & Brahmachari, S. (2020). Cryptocurrency mechanisms for

blockchains: models, characteristics, challenges, and applications. In Elsevier eBooks (pp. 323–348). https://doi.org/10.1016/b978-0-12-819816-2.00013-7

[2] Zhao, H., & Zhang, L. (2021). Financial literacy or investment experience: which is more influential in cryptocurrency investment? International Journal of Bank Marketing, 39(7), 1208-1226. https://doi.org/10.1108/IJBM-11-2020-0552

[3] Amsyar, I., Christopher, E., Dithi, A., Khan, A. N., & Maulana, S. (2020). The Challenge of Cryptocurrency in the Era of the Digital Revolution: A Review of Systematic Literature. Aptisi Transactions on Technopreneurship (ATT), 2(2), 153–159. https://doi.org/10.34306/att.v2i2.96

[4] Delfabbro, P., King, D. L., & Williams, J. N. (2021). The psychology of cryptocurrency trading: Risk and protective factors. Journal of Behavioral Addictions, 10(2), 201–207. https://doi.org/10.1556/2006.2021.00037

[5] Cryptocurrency Scams: Analysis and Perspectives. (2021). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9591634

[6] Livieris, I. E., Kiriakidou, N., Stavroyiannis, S., & Pintelas, P. (2021). An Advanced CNN-LSTM Model for Cryptocurrency Forecasting. Electronics, 10(3), 287. https://www.mdpi.com/2079-9292/10/3/287

[7] Fang, F., Ventre, C., Basios, M., Kanthan, L., Martinez-Rego, D., Wu, F., & Li, L. (2022). Cryptocurrency trading: a comprehensive survey. Financial Innovation, 8(1), 13. https://doi.org/10.1186/s40854-021-00321-6

[8] Fauzi, M. A., Paiman, N., & Othman, Z. (2020). Bitcoin and Cryptocurrency: Challenges, opportunities and future works. The Journal of Asian Finance, Economics and Business, 7(8), 695–704. https://doi.org/10.13106/jafeb.2020.vol7.no8.695

[9] Momtaz, P. P. (2021). The Pricing and Performance of Cryptocurrency. The European Journal of Finance, 27(4-5), 367-380. https://doi.org/10.1080/1351847X.2019.1647259

[10] Astuti, I. D., Rajab, S., & Setiyouji, D. (2022). Cryptocurrency Blockchain Technology in the Digital Revolution Era. Aptisi Transactions on Technopreneurship (ATT), 4(1), 9–15. https://doi.org/10.34306/att.v4i1.216