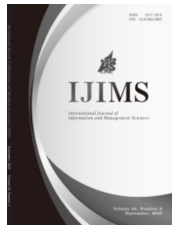
http://ijims.ms.tku.edu.tw/main.php

International Journal of
Information
and
Management
Sciences

International Journal of
Information and Management Sciences

35 (2024), 191-202. DOI:10.6186/IJIMS.202409_35(3).0001



Forecasting Bitcoin Returns via Machine Learning Algorithms with Technical and Economic Indicators

Yen-Sheng Lee[1*] *and Ram Basnet*[2]
[1] Davis School of Business, Colorado Mesa University, U.S.A.
[2] Department of Computer Science and Engineering, Colorado Mesa University, U.S.A.

Keywords

Machine Learning; Bitcoin Returns; Technical Indicators; Economic Indicators

Abstract

Research findings indicate that machine learning models incorporating technical indicators and pattern-based signals derived from Bitcoin's past price information may be informative of variations in Bitcoin price. Our empirical study uses technical and economic indicators to examine the predictability of Bitcoin's daily returns. We apply two machine learning algorithms: support vector machines and gradient-boosting decision trees. The robustness of both algorithmic mechanisms is tested by a K-fold cross-validation method. While findings in the literature yield 50% to 60% accuracy in classifying Bitcoin returns, the average accuracy rate of our predictive models is around 70%, demonstrating the effectiveness of economic indicators and a degree of inefficiency in the Bitcoin market. Our results show that machine learning approaches with technical and economic features can help traders anticipate short-term movements in the Bitcoin market.

1. Introduction

In recent years, the rapid rise in popularity of cryptocurrencies, particularly Bitcoin, has captured the attention of investors and researchers (Tierno, 2023). With its decentralized nature and volatile market behavior, Bitcoin presents a unique challenge and opportunity to apply machine learning techniques to predict price movements. The application of machine learning in forecasting Bitcoin prices has gained significant traction due to its potential to uncover patterns in vast amounts of historical data and inform predictions about future trends. This introduction explores how machine learning algorithms are leveraged to predict Bitcoin prices, the challenges involved, and the potential implications for investors and the broader cryptocurrency ecosystem.

Machine learning algorithms, ranging from traditional statistical models to advanced deep learning techniques, have been employed to forecast Bitcoin prices with varying degrees of

*corresponding author. Email: ylee@coloradomesa.edu

success (Pintelas et al., 2020; Kumar et al., 2023). These algorithms analyze historical price data, market sentiment, trading volumes, and other relevant factors to identify patterns and trends that may influence future price movements. Some common machine-learning approaches for Bitcoin price prediction include time series analysis, regression models, support vector machines, and neural networks.

Regression models, including linear regression, ridge regression, and lasso regression, are employed to identify the relationship between Bitcoin prices and various explanatory variables, such as trading volume, transaction fees, network activity, and macroeconomic indicators (Saheed et al., 2022). These models attempt to estimate the coefficients of these variables to predict future price movements based on the historical relationships to Bitcoin prices.

Support vector machines (SVMs) are powerful supervised learning algorithms for classification and regression tasks, including Bitcoin price prediction (Chen et al., 2020). SVMs seek to identify the optimal hyperplane that separates different classes or predicts continuous values, such as Bitcoin prices, by maximizing the margin between data points and the decision boundary.

Neural networks, particularly deep learning architectures like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, show promise in capturing complex patterns and nonlinear relationships in Bitcoin price data (Kumar et al., 2023). These models are trained using sequential data and historical price trends to make more accurate and dynamic predictions, especially over longer time horizons.

Challenges: Despite advancements in machine learning techniques for Bitcoin price prediction, several challenges remain (Pintelas et al., 2020). The inherent volatility and unpredictability of cryptocurrency markets make it challenging to develop accurate forecasting models that consistently outperform traditional approaches. Moreover, a lack of transparency, regulatory uncertainty, and market manipulation in the cryptocurrency space poses additional challenges for machine learning-based prediction models (Pintelas et al., 2020).

Data quality and feature selection are critical to the performance of machine learning algorithms in predicting Bitcoin prices. The availability of high-quality, reliable data sources and the identification of relevant features that drive price movements are essential for building robust prediction models.

Furthermore, the dynamic nature of cryptocurrency markets, characterized by sudden price spikes, crashes, and regime shifts, presents a moving target for machine learning models. Adapting to changing market conditions and incorporating new information in real time poses a significant challenge for static prediction models trained on historical data.

Implications: Despite these challenges, the application of machine learning in predicting Bitcoin prices is valuable to investors, traders, and policymakers (Tierno, 2023). Accurate price forecasts can help investors make informed decisions, mitigate risks, and capitalize on profitable trading opportunities in cryptocurrency markets. Moreover, reliable prediction models can enhance market efficiency, liquidity, and stability, contributing to the maturation and mainstream adoption of cryptocurrencies.

The application of machine learning to predict Bitcoin prices represents an exciting frontier in financial forecasting. While challenges persist, ongoing research and innovation in this field are likely to yield more accurate and reliable prediction models, with far-reaching implications for the cryptocurrency ecosystem and the broader financial industry.

Our study focuses on the use of machine learning algorithms to forecast the Bitcoin market using technical and economic indicators. The rest of the paper is organized as follows. Section 2 reviews the related literature, which is followed by an overview of the data and methodology in Section 3. Subsequently, Section 4 discusses the results and robustness check. Section 5 concludes our study and points out the direction for future research.

2. Literature Review

This section reviews the financial economics literature on the application of technical analysis and economic indicators to the Bitcoin market. This is followed by a review of the literature on the application of machine learning to cryptocurrencies.

2.1 Technical Analysis and Economic Indicators

One stream of research in financial economics assesses the profitability of technical analysis in cryptocurrency markets. The results of out-of-sample testing tend to support the existence of significant return predictability. Specifically, distinctive technical trading strategies can generate a significant alpha and outperform buy-and-hold positions (Detzel et al., Gerritsen et al., 2020, 2021; Liu, 2019). A study by Svogun and Bazan-Palomino (2022) finds that transaction costs (bubble periods) increase the likelihood of excess return for Bitcoin and Ethereum (Ethereum, Ripple, and Litecoin). Both transaction costs and bubble periods determine the trading profits of technical rules on various cryptocurrencies.

Recent papers scrutinize Bitcoin returns relative to a set of prominent economic factors and validate the viewpoint that Bitcoin has matured from a speculative currency into a mature financial asset. Indeed, Bitcoin has progressed through five distinct phases: Debut, propagation, securitization, liberalization, and solidification (Lee et al., 2022; Vo et al., 2022). Like mature investment instruments, such as stocks and bonds, Bitcoin responds to underlying macroeconomic indicators (e.g., gross domestic product, commodity prices, inflation, interest rates, and volatility). Our two lines of inquiry – technical and economic analysis of Bitcoin historical prices – rely on the assertion that technical and economic determinants are tied to Bitcoin returns.

2.2 Application of Machine Learning to Cryptocurrencies

The three major types of machine learning algorithms are supervised, unsupervised, and reinforcement learning algorithms. Most existing machine learning studies focused on predicting the future value of Bitcoin adopt supervised learning algorithms that learn from an in-sample training dataset (i.e., already-labeled Bitcoin returns, being either positive or negative). This method allows the detection of the relationships between Bitcoin values and the output labels until sufficiently correct labeling results are produced with never-before-seen data (out-of-sample testing dataset). Data scientists deploy various supervised learning algorithms, such as decision trees, support vector machines, and neural networks, for the return prediction of cryptocurrency markets.

One strand of the literature in supervised learning suggests that support vector machines are suitable for the Bitcoin price time series prediction (Ben Hamadou et al., 2023; Souza et al., 2019). For example, Chen et al. (2020) find that support vector machines are superior to statistical methods, with an accuracy of 65.3% for Bitcoin 5-minute interval price prediction. Similarly, Gurrib & Kamalov (2022) find that support vector machines using Bitcoin market price and sentiment information (i.e., news headlines and posts on social media) as input features result in forecast accuracy of 58.5% relative to the direction of the next-day price movement of Bitcoin. The achievement of above 58%-65% prediction accuracy among existing studies implies that the Bitcoin market is inefficient and contradicts the efficient-market hypothesis (Welch & Goyal, 2008). The efficient-market hypothesis states that financial markets are efficient in that the trading prices of financial securities follow a random walk pattern (i.e., the price randomly moves upward and downward 50% of the time, respectively). This theory indicates that the predictability of securities price should converge on 50%.

Although there is no perfect tool that can anticipate the returns in the Bitcoin market with 100% accuracy, a small number of papers investigate whether Bitcoin price is foreseeable using technical indicators that are constructed based on Bitcoin's historical price – open, high, low, close prices and trading volume. These studies find that machine learning models, along with technical analysis, have potential predictive power for Bitcoin price dynamics driven by economic fundamentals (Cocco et al., 2021; Huang et al., 2019)

Most existing studies examine Bitcoin returns from the viewpoints of financial economics research and machine learning frameworks. Instead, we investigate whether the incorporation of economic factors into machine learning models increases the accuracy rate of Bitcoin return prediction.

3. Data and Methodology

3.1 Data

Macroeconomic data was downloaded from Federal Reserve Economic Data (FRED) published by the Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org/), including gross domestic product (*GDP*), an inflation expectation rate (*T5YIFR*), unemployment rate (*UNRATE*), 10-Year Treasury constant maturity rate (*DGS10*), effective Federal Funds rate (*FEDFUNDS*), the spread between 10-Year and 3-Month Treasury constant maturity (*T10Y3M*), investor's sentiment (*VIXCLS*) measured by the market expectation of near-term volatility conveyed by stock index option prices, S&P 500 Index (*S&P 500*), crude oil price (*Oil*), gold price (*Gold*), M1 money stock (*M1*), M2 money stock (*M2*), total assets of the Federal Reserve (*TAFR*), total public debt divided by *GDP* (*TPD/GDP*), St. Louis Fed financial stress index (*SLFFSI*), and Chicago Fed national financial conditions index (*CFNFCI*). These prominent 16 economic variables are strongly correlated with the movement of the financial markets and are expected to be associated with Bitcoin price. Note that monthly or quarterly economic data is interpolated into daily data that matches with the Bitcoin daily price dataset.

Our sample of the Bitcoin price time series was downloaded from the Yahoo@Finance website (https://finance.yahoo.com/) from October 2010 to January 2023. Based on the Bitcoin price time series, 124 technical indicators were computed by using the package from the Technical Analysis library (TA-Lib) implemented in Python 3.8 (Benediktsson et al., 2017). TA-

Lib is extensively used by trading software programmers to perform technical analysis of financial market data. These 124 indicators are categorized into ten groups: momentum, cycle, math operators, math transforms, overlap study, pattern recognition, price transform, statistic functions, volatility, and volume groups. The complete list and detailed description of the 124 indicators can be found on GitHub (https://ta-lib.github.io/ta-lib-python/).

Our dataset consists of 130 input features or independent variables, including the percentage change among 16 economic and 124 technical indicators. The dependent variable, Bitcoin return, is calculated by subtracting the opening price from the closing price (i.e., daily gains or losses) and then dividing by the opening price. As noted earlier, the goal of this study is to predict the range, rather than the level, of the next-day Bitcoin return. We divide the domain of Bitcoin return into eight non-overlapping return ranges, and our model predicts which range the returns will fall into. The eight return ranges comprise four intervals with only positive returns, four intervals with only negative returns, and one impartial interval. The first set of the four positive return intervals are (0.2%, 8%), (8%, 16%), (16%, 24%), and (24%, ∞), labeled as "1", "2, "3", and "4". The second set of four negative return intervals are (−100%, −24%), (−24%, −16%), (−16%, −8%), and (−8%, −0.2%), labeled as "5", "6, "7", and "8". The impartial interval registers returns between −0.20% and 0.20% and is labeled as "0." Excluding missing values, our full dataset has 4,107 observations, and we split 70% and 30% of our full dataset into the training and test samples. As a result, these two subsamples include 2,875 and 1,232 observations, respectively. The following methodology section provides a high-level explanation of the machine learning algorithms explored in this study.

3.2 Methodology

Two techniques of classification, support vector machines, and gradient-boosting decision trees, are popular for financial market prediction. Support vector machines (Evgeniou & Pontil, 2001) are a classifier that maps input vectors into a high-dimensional feature space (also known as a hyperplane) that separates data points into different classes. Support vector machines select a hyperplane with a maximized margin that represents the greatest distance between data points of those classes. Gradient-boosting decision trees incorporate a regularizing gradient-boosting framework into a decision tree classifier (Chen & Guestrin, 2016). Gradient-boosting decision trees frequently achieve higher accuracy than a single decision tree but forgo the interpretability of decision trees. We implemented both classification algorithms using *Scikit-learn* functions in Python 3.8 (Pedregosa et al. 2011). The accuracy of our predictive models can be derived in Equation (3.1) as follows:

$$Accuracy = (tp + tn)/(tp + tn + fp + fn) \tag{3.1}$$

where *tp* and *tn* stand for a true (that is, the real label is equal to the predicted label) positive and negative, and *fp* and *fn* denote a false (that is, the real label is not equal to the predicted label) positive and negative, respectively.

Mathematical explanations for the two classifiers are beyond the scope of our research, and readers interested in the technical details are encouraged to follow the citations mentioned above.

4. Results and Robustness Check

Section 4.1 outlines the results of our empirical analysis of the out-of-sample predictive power. Section 4.2 deploys the *K*-fold cross-validation method as a robustness check to test the validity of our proposed models.

4.1 Out-of-sample predictive power for Bitcoin return

Fig. 1 plots the out-of-sample accuracy of Bitcoin return prediction over the number of input variables (features) for support vector machines. As can be seen from the figure, accuracy increases quickly and plateaus around 75% over the top 20 features. With 40 features, the accuracy reaches around 80% but begins to slowly decline to 75% among all features. Similar to Fig. 1, Fig. 2 plots the out-of-sample accuracy of gradient-boosting decision trees and shows a comparable trend, although accuracy levels off at 80% with the best 20 features and remains the same over the remainder of the features. Because the leading 20 features come close to attaining maximum accuracy, Table 1 presents the features shared across both algorithms, suggesting that the most important features are consistent between our proposed models. On top of these features are *DGS* 10, S&P 500, *M1*, and *TPD*/*GDP* economic indicators. Our finding suggests that the inclusion of these four economic features enhances the rigor and accuracy of our models. We also use a *K*-fold cross-validation method to verify the results of our test sample, with details included in the next section.

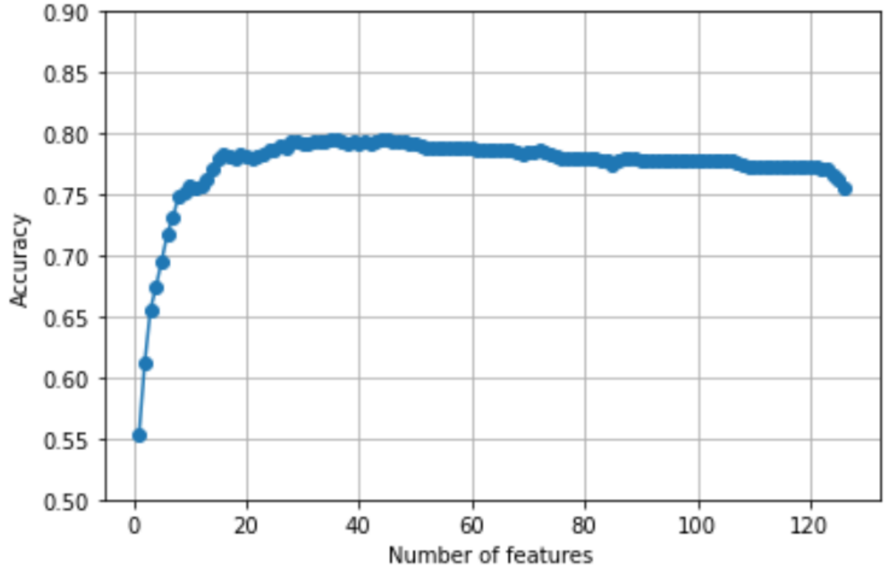


Fig. 1 *Accuracy of Bitcoin Return Prediction of Support Vector Machines*

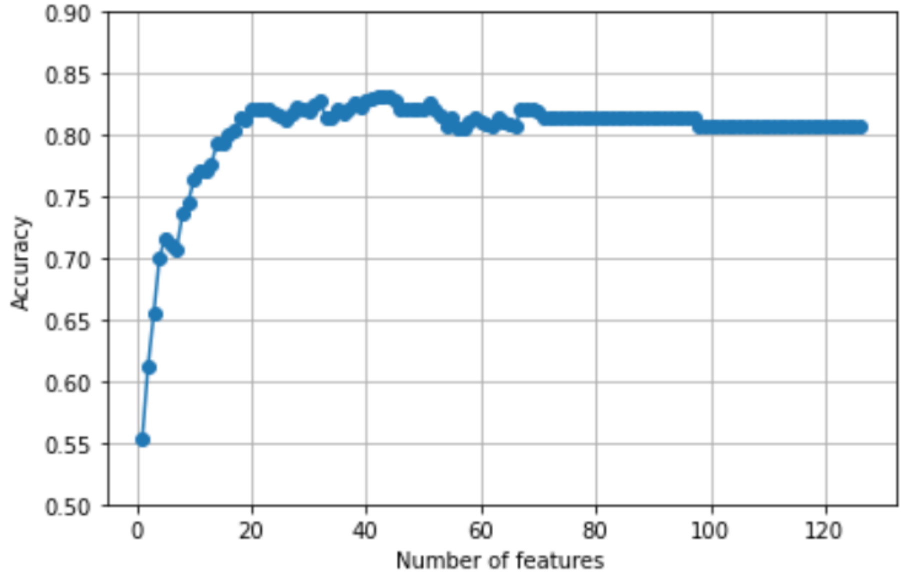


Fig. 2 *Accuracy of Bitcoin Return Prediction of Gradient-boosting Decision Trees*

Table 1 *Top 20 Features of Support Vector Machines and Gradient-boosting Decision Trees*

Economic and technical indicators	Description
*DGS 10**	10-Year Treasury constant maturity rate
*S&P 500**	S&P 500 Index
*M1**	*M1* money stock
*TPD/GDP**	Total public debt divided by *GDP*
MAX	Highest value over a specified period (day)
MIN	Lowest value over a specified period (day)
TANH	Vector hyperbolic tangent
CDL3STARSINSOUTH	Three stars in the south
CDLBELTHOLD	Belt-hold
CDLBREAKAWAY	Breakaway
CDLDOJISTAR	Doji star
CDLEVENINGDOJISTAR	Evening doji star
CDLHANGINGMAN	Hanging man
CDLHARAMICROSS	Harami cross pattern
CDLHIKKAKE	Hikkake pattern
CDLHIKKAKEMOD	Modified hikkake pattern
CDLMARUBOZU	Marubozu
CDLSPINNINGTOP	Spinning top
CDLSTALLEDPATTERN	Stalled pattern
TRANGE	True range

*Note: * denotes* economic indicators.

4.2 Robustness check

The main purpose of the robustness check is to verify whether prediction bias exists in our test sample due to sampling bias. K-fold cross-validation, a statistical method to validate predictive models, splits the full sample into K subsamples, each of which serves as the validation dataset. We chose the value of K to be 10, which is common in the data science field, and thus divided the full sample of 4,107 observations into 10 validation datasets, each with 410 observations. We compute the accuracy of Bitcoin return prediction on each of 10 validation datasets by means of the foremost 20 features listed in Table 1. The results of K-fold cross-validation are summarized in Table 2.

Table 2 *Results of K-fold Cross-validation*

# of validation dataset	Accuracy of support vector machines	Accuracy of gradient-boosting decision trees
1	77.8%	81.9%
2	74.3%	76.0%
3	74.3%	78.8%
4	71.9%	78.8%
5	76.0%	74.6%
6	76.7%	78.7%
7	75.3%	76.3%
8	74.9%	79.4%
9	75.3%	75.3%
10	75.6%	75.6%
Average	75.2%	77.6%

Across 10 validation datasets, the averages and ranges of accuracy are 75.2% (77.6%) and 1.5% (2.2%) for support vector machines (gradient-boosting decision trees). The results of K-fold cross-validation are similar to those of out-of-sample accuracy in Section 4.1 and suggest there is little or no prediction bias in our proposed models.

5. Conclusion

Our study responds to a growing interest in the Bitcoin market and sheds light on the directional movement of Bitcoin price by synthesizing theory and methods from the fields of financial economics and machine learning. The findings of our two machine learning techniques suggest that four economic indicators (*DGS 10*, *S&P 500*, *M1*, and *TPD*/*GDP*) are discriminating features in forecasting the daily return in the Bitcoin market. Our contributions to the existing literature are twofold. First, our analytical models are equipped with economic-based features and generate a higher accuracy of around 75% compared to the 50% to 60% accuracy indicated elsewhere in the literature. We also find that economic-based features play a more important role in technical indicators. Second, our empirical evidence suggests that some degree of inefficiency is present in the Bitcoin market, in contrast to the efficient-market hypothesis, which stipulates that financial markets are efficient and that accuracy in forecasting security returns should not exceed 50%. Even though our results indicate some inefficiency in Bitcoin price movement, whether insufficiency is the driver of abnormal returns warrants further investigation.

An interesting direction for future research would be to incorporate more macroeconomic factors and analyze the effect of other macroeconomic factors. Researchers can classify Bitcoin returns into different ranges, i.e., (0.5%, 5%), (5%, 10%), (10%, 20%), and (20%, ∞), and examine the impact of various return ranges on accuracy. Future studies could also apply artificial intelligence models with technical and economic factors to other types of cryptocurrencies, e.g., Ethereum, XRP, Litecoin, and Bitcoin cash.

References

Ben Hamadou, F., Mezghani T., Zouari, R., & Boujelbene-Abbes, M. (2023). Forecasting Bitcoin returns using machine learning algorithms: impact of investor sentiment. *EuroMed Journal of Business*, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/EMJB-03-2023-0086

Benediktsson, J., & Other Contributors. (2017). TA-lib Indicators. https://ta-lib.github.io/ta-lib-python/doc_index.html

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* 785–794. https://doi.org/10.1145/2939672.2939785

Chen, Z., Li, C., & Sun, W. (2020). Bitcoin price prediction using machine learning: An approach to sample dimension engineering. *Journal of Computational and Applied Mathematics*, *365*, 107-137. https://doi.org/10.1016/j.cam.2019.112395

Cocco, L., Tonelli, R., & Marchesi, M. (2021). Predictions of bitcoin prices through machine learning based frameworks. *PeerJ Computer Science* 7:e413. https://doi.org/10.7717/peerj-cs.413

Detzel, A., Liu, H., Strauss, J., Zhou, G., & Zhu, Y. (2021). Learning and predictability via technical analysis: Evidence from bitcoin and stocks with hard-to-value fundamentals. *Financial Management*, *50*(1), 107-137. https://doi.org/10.1111/fima.12310

Evgeniou, T., & Pontil, M. (2001). Support vector machines: theory and applications. *Machine Learning and its Applications*, *2049*, 249–257. https://link.springer.com/chapter/10.1007/3-540-44673-7_12

Gerritsen, D., Bouri, E., Ramezanifar, E., & Roubaud, D. (2020). The profitability of technical trading rules in the Bitcoin market. *Finance Research Letters*, *34*, 101263. https://doi.org/10.1016/j.frl.2019.08.011

Gurrib, I., & Kamalov, F. (2022). Predicting bitcoin price movements using sentiment analysis: a machine learning approach. *Studies in Economics and Finance*, *39*(3), 374-364. https://doi.org/10.1108/SEF-07-2021-0293

Huang, J., Huang, W., & Ni, J. (2019). Predicting bitcoin returns using high-dimensional technical indicators. *The Journal of Finance and Data Science*, *5*(3), 140-155. https://doi.org/10.1016/j.jfds.2018.10.001

Kumar, V., Santhi, S., Shanthi, K., & Gokila, M. (2023). Cryptocurrency price prediction using LSTM and recurrent neural networks. *2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India, 1-5. doi:10.1109/ICAAIC56838.2023.10141048

Lee, Y., Vo, A., & Chapman, T. (2022). Examining the maturity of bitcoin price through a catastrophic Event: The case of structural break analysis during the COVID-19 pandemic. *Finance Research Letters*, *49*, 103165. https://doi.org/10.1016/j.frl.2022.103165

Liu, L., (2019). Are bitcon returns predictable? Evidence from technical indicators. *Physica A: Statistical Mechanics and its Applications*, *533*, 121950. https://doi.org/10.1016/j.physa.2019.121950

Pedregosa F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ...& Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Pintelas, E., Livieris, E., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. *In proceedings of the IFIP Advances in Information and Communication Technology, 584,* 99-110. https://link.springer.com/chapter/10.1007/978-3-030-49186-4_9

Saheed, Y., Ayobami, R., & Orje-Ishegh, T. (2022). A comparative study of regression analysis for modelling and prediction of Bitcoin price. *In: Misra, S., Kumar Tyagi, A. (eds) Blockchain Applications in the Smart Era. EAI/Springer Innovations in Communication and Computing. Springer, Cham.* https://doi.org/10.1007/978-3-030-89546-4_10

Souza, M., Amudhaf, F., Henruque, B., Negredo, A., Ramos, D., Sobreiro, V., & Kimura, H. (2019). Can artificial intelligence enhance the Bitcoin bonanza? *The Journal of Finance and Data Science*, *5*(2), 83-98. https://doi.org/10.1016/j.jfds.2019.01.002

Svogun, D., & Bazan-Palomino, W. (2022). Technical analysis in cryptocurrency markets: Do transaction costs and bubbles matter? *Journal of International Financial Markets, Institutions and Money*, *79*, 101601. https://doi.org/10.1016/j.intfin.2022.101601

Tierno, P. (2023, February 15). Cryptocurrency: Selected Policy Issues. *Congressional Research Service.* https://crsreports.congress.gov/product/pdf/R/R47425. Last accessed on Feb 19, 2024.

Vo, A., Chapman, T., & Lee, Y. (2022). Examining bitcoin and economic determinants: An evolutionary perspective. *Journal of Computer Information Systems*, *62*(3), 572-586. https://doi.org/10.1080/08874417.2020.1865851

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455-1508. https://doi.org/10.1093/rfs/hhm014

Yen-Sheng Lee
Davis School of Business, Colorado Mesa University, USA.
E-mail address: ylee@coloradomesa.edu
Major area: Finance

Ram Basnet
Department of Computer Science and Engineering, Colorado Mesa University, USA.
E-mail address: rbasnet@coloradomesa.edu
Major area: Computer Science

(Received May 2024; accepted September 2024)