International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-008
DOI: https://doi.org/10.5281/zenodo.19124046

PREDICTIVE ANALYTICS FOR RISK MANAGEMENT IN CRYPTOCURRENCY MARKETS

#1GITHA SRUTHI, ***Assistant Professor,***

#2K.SAVITHA, ***Assistant Professor,***

Department of Computer Science & Engineering,

Mother Theressa College of Engineering & Technology, Peddapalli, Telangana.

ABSTRACT: This study investigates the effectiveness of machine learning models for predicting cryptocurrency price movements and supporting trading decisions. The proposed models forecast daily binary relative price changes for the top 100 cryptocurrencies. Experimental results show average prediction accuracies between 52.9% and 54.1%, indicating statistically meaningful predictive performance. When only high-confidence predictions (top 10%) are considered, accuracy improves to a range of 57.5% to 59.5%.To evaluate practical applicability, long–short portfolio strategies were implemented using ensemble models based on Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) networks. After accounting for transaction costs, the out-of-sample annualized Sharpe ratios reached 3.23 for the GRU ensemble and 3.12 for the LSTM ensemble, significantly outperforming the buy-and-hold benchmark Sharpe ratio of 1.33. These findings suggest the existence of predictable patterns in cryptocurrency markets and highlight potential inefficiencies despite possible arbitrage limitations.

Keywords:*Machinelearning;GRU;LSTM;Neuralnetwork;Randomforest;Gradientboosting;Temporalconvolutionalneuralnetwork*

1. INTRODUCTION

In 2008, Nakamoto formally unveiled Bitcoin, a decentralized digital currency. Numerous additional cryptocurrencies with distinct technological characteristics and potential applications beyond Bitcoin surfaced in the years that followed. The value of various digital currencies has fluctuated over the past ten years due to the exponential growth of the cryptocurrency market. Due of space constraints, the user's text cannot be edited academically. Regarding the usefulness of cryptocurrencies like Bitcoin, market participants are divided. They input two, three, and five digits.

In this research, auto-regressive statistical techniques are frequently employed to clearly demonstrate non-linear connections. Because machine learning algorithms understand the dynamic interplay between causes and objectives, they have been used for predicting purposes in the Bitcoin and cryptocurrency markets.The two numbers that are mentioned are eight and nine. Notably, these methods can exploit intricate interactions between variables and reveal elements in high-dimensional domains, which are rarely visible in market performance research.

Finding the most effective machine learning models for forecasting the direction of the financial markets was the aim of this research. As a result, we are motivated to respond to this crucial query. We used six machine learning classifiers to forecast the daily performance of the top 100 cryptocurrencies in order to address this research issue. For each model, we use the out-of-sample estimates to develop a long-short trading strategy.

We then examine the results of each transaction. Each of the five historical eras is given 400 days out of the 800 total. This research has two noteworthy additions: First, it should be noted that all of the models that are employed produce predictions for the bitcoin market that are statistically sound, demonstrating the potency of machine learning as a predictor. Our findings demonstrate that recurrent neural networks or tree-based ensembles outperform other approaches when comparing the daily prices of multiple cryptocurrencies. The long-short portfolio strategy outperforms the market average despite transaction expenses. This implies that the Bitcoin market may present statistical arbitrage opportunities.

2. RELATED WORK

Fischer et al. (2018) investigated whether machine learning predictions could improve statistical arbitrage in the bitcoin market by examining data collected between June and September. Based on the temporal distribution of historical returns from the previous day, the research employed a logistic regression model and a random forest classifier to forecast the relative performance of the top 40 cryptocurrencies over the course of the next 120 minutes. Given that an out-of-sample long-short trading strategy based on the researchers' model projections yielded a daily return of 7.1 basis points, the findings might lead to a decrease in the efficiency of the bitcoin market.

Fil and Kristoufek (2010) reexamined the concept of pairs trading as it relates to the cryptocurrency market, taking into account the long-term stability of many cryptocurrency pairings. This research looked at daily, hourly, and 5-minute trade activity from January 2018 to September 2019. The enormous volume of bitcoin transactions has shown that pair trading is viable (Fil and Kristoufek, 2018). Certain aspects of the market, including transaction costs, have a significant impact on the trading strategy's effectiveness. In order to evaluate the effectiveness of deep reinforcement learning in bitcoin trading, Betancourt and Chen (2011) used data from August 2017 to November 2020. The agents of the proposed system make daily trading decisions based on 20 days of data on a cryptocurrency's volume, market capitalization, price, and price history. The method created by Betancourt and Chen (2011, IEEE) may facilitate Bitcoin transactions. McNally et al. evaluated three methods for predicting daily changes in the Bitcoin binary market in 2012. Other techniques that were examined included Elman recurrent neural networks, extended short-term neural networks, and autoregressive integrated moving averages. Using data from 2013 to 2016, the extended short-term neural network outperformed its competitors with a model accuracy of 52.78%.

Dutta et al. (2013) investigate many neural network algorithms that employ a range of technical, blockchain-based, asset-based, and interest-based components in order to forecast the daily value of Bitcoin. Statistics from 2010 to 2019 indicated that the most successful method was to use a gated recurrent unit with recurrent dropout.

Chen et al. (2014) used a combination of machine learning and linear statistical techniques to anticipate the 5-minute and daily Bitcoin markets. The period of the research was 2017–2019. For daily predictions, statistical approaches perform better than machine learning techniques. Alessandretti et al. (2015) calculated the daily returns of 1,681 distinct cryptocurrencies using sophisticated short-term neural network algorithms and gradient boosting classifiers. The authors use empirical data gathered between 2015 and 2018 to demonstrate that forecast-based portfolio strategies outperformed a baseline strategy. Lahmiri and Bekiros (2016) examined two neural network models. While one network trained using long short-term memory, the other network employed extended regression. The goal of the research was to create a model that could predict the future value evolution of digital currencies such as Bitcoin, Ripple, and Digital Cash. The researchers' data sets encompassed everything from ancient times to last October. They demonstrate that neural networks trained for short-term memory storage outperform generalist regression networks.

3. METHODOLOGY

The research is based on the four elements identified by Fischer et al.8 and Fischer Krauss17. We search the internet for relevant information in this first stage. Using the original price data, we construct features and goals for forecasting the currency's future returns. Divide the dataset in half equally. Since the first area is only for testing, market elements should be eliminated; but, since the second sector is for research, they should be included. We can now perform backtests with ease. Each deployed model must be trained and improved throughout predefined time periods in order to complete the trading prediction simulation process. Specifics The research used daily closing prices and market capitalization data from the CoinGecko (CG) API. The USD values data collection period runs from February 8, 2018, to May 15, 2022. details regarding the cryptocurrency market value. To guarantee that our research of the investing environment is objective and unaffected by survivorship bias, we remove the top 100 cryptoassets from the training dataset on the first trading day. This technique reduces "look-ahead bias" in coin universe creation while ensuring that all coins have enough training data. Since the value of these stablecoins is either fixed or closely linked to a fiat currency, such as the US dollar, they are not eligible. Ten more coins were excluded from the analysis due to errors or missing numbers in the data sources. The complete list of prohibited coins can be seen in Appendix B.1. The CG API currently provides daily USD market capitalization information for the top 1,750 currencies as of June 8, 2022. This is done using the known amount and the asset's market value. On the first day of the training set, all cryptocurrencies are sorted by market capitalization before each session starts. As a result, creating bitcoin asset portfolios is made simpler. Information on the amount of money in different coins

The market price data from CoinGecko is used to compute the return. The prices displayed on the CG platform are the mean of all cryptocurrency pairings—both fiat money and cryptocurrency—that have been documented. This information was gathered from a large number of interactions. The total number of transactions determines the pricing. The result of 19 indicates that these figures accurately reflect the current state of the bitcoin industry, despite the fact that they represent unsold prices. The incorporation of prices from other exchange platforms has no effect on the liquidity of the bitcoin market, according to the author's observations. Since exchanges are active all day, it is simple to use the market price at midnight UTC to affect the closing price of Bitcoin. Every day at midnight UTC, the CG API is updated with the pricing information for the current day. Simply raise the daily quote time series by one day to obtain the results from the day before. To compute the return, the market price of coin c on day t, denoted by rm,c, must be known.

The measurement is made on day t+1 at midnight UTC. Add Coin C's closing price on day t to determine how much it has increased in value over the previous m days. When m=1, you may use this method to determine the asset's daily earnings; when m>1, you can use it to determine the total returns for the previous m days. To get a 20% return, you don't have to

International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-008
DOI: https://doi.org/10.5281/zenodo.19124046

play any games. One can estimate excess returns by applying the secondary market interest rate on three-month Treasury Bills. T-bills are a type of short-term loan issued by the US Treasury.

The organization of the research period and the division of the train, validation, and test sets are depicted in Figure 1.

The state government has a three-month term. The annual interest rate must be converted into daily returns in order to compute risk-adjusted return measures like the Sortino and Sharpe ratios. With an average of 3.9 x 10^6 throughout the course of the research, the risk-free rate, which varies daily between 2.4 x 10^7 and 2.8 x 10^5, is continuously low, as demonstrated by the aforementioned T-bill rate.

Software and hardware

This project uses Python 3.9 to handle data gathering, processing, and analysis. Pandas22 handles data processing, whereas Numpy21 handles feature creation. Many use the Scikit-learn library for the development and training of classic machine learning models. Keras is used in conjunction with the TensorFlow backend to construct deep learning models. Intel Core i5-8400 CPUs running at 2.8 GHz are used to train the models.

Data split

The five research periods (SPs) in this research add up to 800 trading days, which is the prediction goal. Each model uses data from the previous three months to generate forecasts. Thus, information from the ninety days preceding the first trading day is used to compute the SP for each asset. The research's time frame is depicted in a triangle in Figure 1. This framework consists of a training set, validation set, and out-of-sample test set. We collect data for 500 days in order to train the model. We utilize the 150-day validation set for hyperparameter adjustment.

To evaluate the model's performance, a 150-day out-of-sample test set is employed. Table 1 provides a detailed analysis of the distribution of each research era across the three data levels.

The formation process includes both training and validation. The models are trained in the

International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-008
DOI: https://doi.org/10.5281/zenodo.19124046

training phase.

Following validation, the validation findings are used to identify the optimal hyperparameters. During each research session, the testing unit is in charge of delivering simulations and actual trade tests. Given that five sets of exams will be given in quick succession, the length of the testing period will dictate the adjustments that must be made to the research sessions. Recalibrating researchers' models on a regular basis can help mitigate concept drift caused by market fluctuations during multiple research sessions.

Features

The fact that every model uses the same training data presents a challenge for binary classification. The objective is to determine which coin will perform better than the cross-sectional average the next day. This conclusion was reached using solely pricing data from the previous ninety days. We examine the returns on all coins from three months prior to trading in order to compute the characteristics for every model.

Table 1Shows the research lengths for the test, validation, and training sets along with the relevant date ranges.

SP No	Training Set	Validation Set	Test Set
1	2018-07-16 - 2019-11-27	2019-11-28 - 2020-04-25	2020-04-26 - 2020-09-22
2	2018-12-13 - 2020-04-25	2020-04-26 - 2020-09-22	2020-09-23 - 2021-02-19
3	2019-05-12 - 2020-09-22	2020-09-23 - 2021-02-19	2021-02-20 - 2021-07-19
4	2019-10-09 - 2021-02-19	2021-02-20 - 2021-07-19	2021-07-20 - 2021-12-16
5	2020-03-07 - 2021-07-19	2021-07-20 - 2021-12-16	2021-12-17 - 2022-05-15

Classifiers with and without memory features are heavily used in the inquiry. Owing to their distinct features, these classifiers are developed separately. Each of these three deep learning models—LSTM, GRU, and TCN—can generate 90-character daily return patterns with varying memory capacities. It is usual procedure to divide the daily mean by the standard deviation when standardizing a training set. Logistic regression (LR) and tree-based classifiers require historical data due to memory limitations.

It is essential to keep a constant stream of input sequences with the most precise labeling. We accomplish this by creating intersecting 90-degree circles, which advances the calendar by one day. A collection of input sequences with the proper labels for use with deep learning methods is provided below.

Figure 2.In order to overcome the limitations of memory-free models like GBC, RF, and LR when working with temporal input data, we aggregate across ever longer time intervals to create time-lagged features. We have included the work of Takeuchi and Lee (26), as well as Krauss et al. (27), based on our findings. The intervals are evaluated with a 10-day increment for m values ranging from 1 to 90. Since we initially compute the daily fluctuations for the

International Journal of Technology, Leadership and Sciences (IJTLS)
ISSN: 3107-9083 www.ijtls.com Paper ID: TLS-26/V2/I2-008
DOI: https://doi.org/10.5281/zenodo.19124046

first 20 days, each sample has a total of 27 characteristics. Equation (1) can be used to determine the effects of different time intervals. Return characteristics and target labels were constructed using logistic regression and tree-based techniques, as illustrated in Figure 3. For all cryptocurrencies and research periods, we create 50,000 training samples, 15,000 validation samples, and 15,000 test samples using both approaches.

Targets

The main goal of the binary forecast challenge is to ascertain whether a single coin will perform better than the cross-sectional median within a day after building a portfolio. The returns for every cryptocurrency are listed in descending order at the end of each trading day. Coins with values below the cross-sectional median are given a value of one under this method, whereas coins with values above the median are given a value of zero. Coin C is currently categorized as follows: The user's text "yc" is insufficiently detailed to qualify as an academic article.

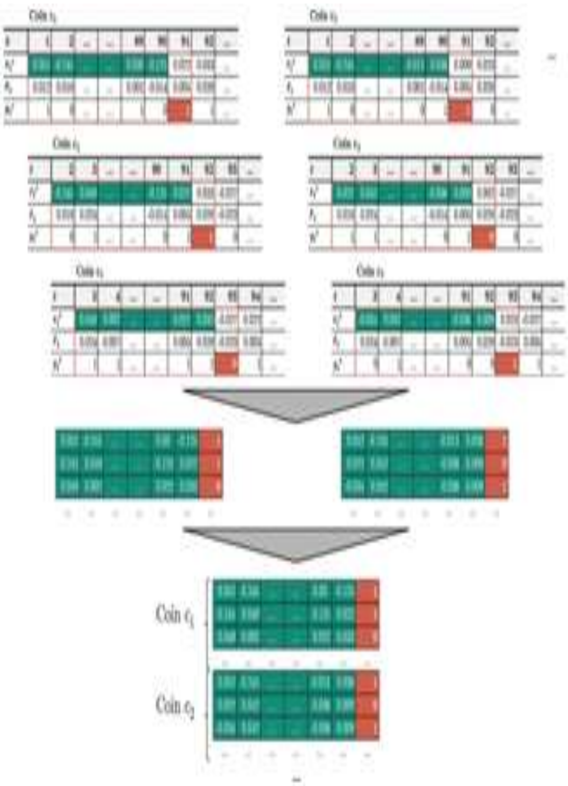

Fig. 2. When a model has storage, it can remember the sequences of features it has created and the labels it has assigned to those sequences.

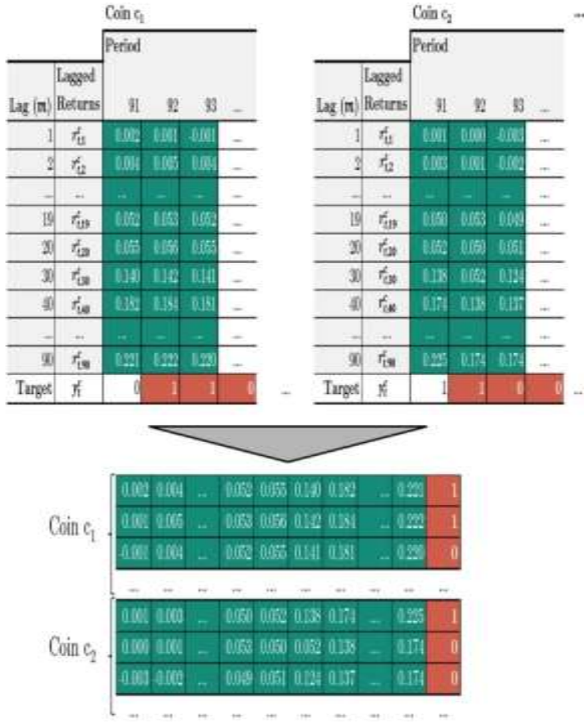

Fig. 3. The development of tree-based and logistic regression models relies heavily on the creation of feature sets and target labels.

4. CONCLUSION

Researchers are trying to guess how the prices of the 100 most valuable cryptocurrencies will change every day by using different machine learning techniques. Each model's anticipated outcomes are shown to be statistically sound in this context. The accuracy of all cryptocurrencies ranges from 52.9% to 54.1% on average; the precise percentage varies depending on the model. We used a sample of forecasts that encompassed the top 10% of each class's highest model confidences during the course of the day in order to calculate the accuracy results. The range of the results was 57.5% to 59.5%. The LSTM and GRU ensemble models used in the long-short portfolio strategy have annual out-of-sample Sharpe ratios of 3.12 and 3.23, respectively, after transaction costs are taken into consideration. A Sharpe ratio of 1.33 indicates the acquisition and upkeep of a benchmark market portfolio. According to this research, arbitrage-related limitations prevented the weak form from performing as well in the bitcoin market. A conclusive evaluation of the effects of these limitations is not possible, yet.

REFERENCES

1. Nakamoto S. Bitcoin: a peer-to-peer electronic cash system. Working Paper URL: https://bitcoin.org/bitcoin.pdf; 2008.

2. Coinmarketcap. Coinmarketcap; 2022. URL: https://coinmarketcap.com/. Accessed July 21, 2022. https://coinmarketcap.com/.
3. Kristoufek L, Vosvrda M. Cryptocurrencies market efficiency ranking: not so straightforward. Phys Stat Mech Appl. 2019;531, 120853.
4. Le Tran V, Leirvik T. Efficiency in the markets of crypto-currencies. Finance Res Lett. 2020;35, 101382.
5. Kakinaka S, Umeno K. Cryptocurrency market efficiency in short-and long-term horizons during covid-19: an asymmetric multifractal analysisapproach. Finance Res Lett. 2022;46, 102319.
6. Gu S, Kelly B, Xiu D. Empirical asset pricing via machine learning. Rev Financ Stud. 2020;33:2223–2273. https://doi.org/10.1093/rfs/hhaa009.
7. Huang JZ, Huang W, Ni J. Predicting bitcoin returns using high-dimensional technical indicators. J Finance Data Sci. 2019;5:140–155.
8. Fischer T, Krauss C, Deinert A. Statistical arbitrage in cryptocurrency markets. J Risk Financ Manag. 2019;12:31. https://doi.org/10.3390/
9. jrfm12010031.
10. Jaquart P, Dann D, Weinhardt C. Short-term bitcoin market prediction via machine learning. J Finance Data Sci. 2021;7:45–66.
11. Fil M, Kristoufek L. Pairs trading in cryptocurrency markets. IEEE Access. 2020;8:172644–172651. https://doi.org/10.1109/ACCESS.2020.3024619.
12. Betancourt C, Chen WH. Reinforcement learning with self-attention networks for cryptocurrency trading. Appl Sci. 2021;11:7377. https://doi.org/10.3390/app11167377.
13. McNally S, Roche J, Caton S. Predicting the price of bitcoin using machine learning. In: Proceedings of 2018 Euromicro International Conference on Parallel, Distributed, and Network-Based Processing. 2018:339–343. https://doi.org/10.1109/PDP2018.2018.00060.
14. Dutta A, Kumar S, Basu M. A gated recurrent unit approach to bitcoin price prediction. J Risk Financ Manag. 2020;13. https://doi.org/10.3390/
15. jrfm13020023. Article 23.
16. Chen Z, Li C, Sun W. Bitcoin price prediction using machine learning: an approach to sample dimension engineering. J Comput Appl Math. 2020;365. https://doi.org/10.1016/j.cam.2019.112395. Article 112395.
17. Alessandretti L, ElBahrawy A, Aiello LM, Baronchelli A. Anticipating Cryptocurrency

Prices Using Machine Learning Complexity; 2018. https://doi.org/10.1155/2018/8983590.

18. Lahmiri S, Bekiros S. Cryptocurrency forecasting with deep learning chaotic neural networks. Chaos, Solit Fractals. 2019;118:35–40.
19. Fischer T, Krauss C. Deep learning with long short-term memory networks for financial market predictions. Eur J Oper Res. 2018;270:654–669. https://doi.org/10.1016/j.ejor.2017.11.054.
20. CoinGecko. Methodology; 2022. https://www.coingecko.com/en/methodology. Accessed July 21, 2022.
21. Vidal-Toma´s D. Which cryptocurrency data sources should scholars use? Int Rev Financ Anal. 2022;81, 102061.
22. Board of Governors of the Federal Reserve System. 3-Month treasury bill secondary market rate. Discount Basis [DTB3]. URL: https://fred.stlouisfed.org/series/DTB3; 2022.