

1. **Journal:** PeerJ Computer Science
2. **Title:** "Predictive performance and robustness of machine learning models in digital currency trading"
3. **Manuscript ID:** CS-2024:07:103313:0:1:NEW 19 Jul 2024
4. **Overall:**
 - a. The main aim of this paper is to analyze machine learning classification models and technical indicators for algorithmic trading in the volatile cryptocurrency market. The study aims to identify practical approaches for informed decision-making and profitable strategy development.
5. **Weaknesses:**
 - a. The authors wrote (lines 72–74), "This paper tries to compare ML-based classification algorithms for the best algorithmic trading in the cryptocurrency market, deriving some insight into the efficacy of different algorithms in securing profitable transactions." In reality, they only evaluated the validity of forecasts from these ML models, not referring to the performance of algorithmic trading based on the forecasts from these models. Such an example of the possibility of applying forecasts from various econometric or ML models in algorithmic investment strategies can be found in Vo Nguyen, Ślepaczuk R., 2022, Applying Hybrid ARIMA-SGARCH in Algorithmic Investment Strategies on S&P500 Index, Entropy, 24(2), 158, <https://doi.org/10.3390/e24020158>
 - b. The problem is that the evaluation of forecast performance is done incorrectly, as described further in this review.
 - c. To properly refer to the potential comparison of the efficiency of various Algo trading strategies based on the forecasts from ML models, the authors would have to use the proper architecture of testing and calculate the adequate performance metrics for the investment based on the buy/sell signals generated based on these forecasts. An example of such calculation of all important performance metrics based on adequately prepared equity lines with some statistical tests verifying the significance of obtained results, can be found in Bui Q., Ślepaczuk R., 2021, Applying Hurst Exponent in Pair Trading Strategies on Nasdaq 100 index, Physica A: Statistical Mechanics and its Applications 592, p. 126784, ISSN = 0378-4371, <https://doi.org/10.1016/j.physa.2021.126784>
 - d. The ML models should be trained with the proper loss function if they are going to be utilized in algorithmic trading strategies. An example of discussion of the validity of the modified optimization criterion (loss function) used in machine learning problems in order to more adequately forecast asset prices can be found in Michańków J., Sakowski P., Ślepaczuk R., 2023, Mean Absolute Directional Loss as a New Loss Function for Machine Learning Problems in Algorithmic Investment Strategies, Journal of Computational Science, <https://doi.org/10.1016/j.jocs.2024.102375>
 - e. very limited description of data and features used in ML models with special attention to the influence of the content of the input layer on the final results of tested models, like for example in Nguyen T.T.G., Ślepaczuk R., 2022, *The efficiency of various types of input layers of LSTM model in investment strategies on S&P500 index*, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 29/2022 (405), https://www.wne.uw.edu.pl/application/files/8316/7083/8777/WNE_WP405.pdf,
 - f. The hyperparameter tuning section is written only in the theoretical sense. The reader does not know which parameters were selected, on which exact period they were tuned, or what were the parameter ranges in the process of their tuning. Moreover, the reader would like to know if the authors used a grid or random search approach. A good example of such a process can be found in Wysocki M., Ślepaczuk R., 2022, *Artificial Neural Networks Performance in WIG20 Index Options Pricing*, Entropy, 24(1), 35, <https://doi.org/10.3390/e24010035>
 - g. No robustness analysis for all parameters, data period, and frequency of data assumed at the beginning of the research. An example of such sensitivity analysis for ML models can be found in: Kijewski M., Ślepaczuk R., Wysocki M., 2024, Predicting prices of S&P500 index using classical methods and recurrent neural networks, Proceedings of the 32nd International Conference on Information Systems Development (ISD 2024), <https://aisel.aisnet.org/isd2014/proceedings2024/datascience/34/>, or in Zenkova M., Ślepaczuk R., 2018, *Robustness of Support Vector Machines in Algorithmic Trading on Cryptocurrency Market*, Central European Economic Journal, 5(1), pp. 186-205, <https://content.sciendo.com/view/journals/ceej/5/52/article-p186.xml>
6. **Title of the paper:**
 - a. Is misleading because it misinforms the reader about the actual content of the article because the presented article does not contain the robustness analysis nor does it refer to the usage of these predictions in algotrading.
7. **Abstract.**
 - a. Please ensure that the abstract covers the following information, of course, in short form:
 - i. The main aim of the research
 - ii. Why the topic is important -> **should be added**
 - iii. The description of the problem
 - iv. The novelty and the contribution/importance of this research -> **should be added.**
 - v. The description of the methodology
 - vi. The description of the data and the frame of the research -> **should be added.**
 - vii. The main results -> **should be added.**
8. **Introduction:**
 - a. Please check if this section covers the following parts:

- i. The introduction to the topic and its main aim
 - ii. The contribution of this research, with a special focus on its novelty -> **should be added.**
 - iii. A description of the methodology and data used -> **should be added.**
 - iv. Hypotheses or research questions stated explicitly in this paper -> **should be added.**
 - v. The structure of this paper -> **should be added.**
- 9. **Literature review:**
 - a. There is no specific literature review section in this paper.
 - b. Such a section should end with a summary of conclusions from previous research, which should state the rational basis for undertaking the new topic, which is the main aim of this research.
 - c. The literature cited in this paper is quite broad and covers many important and recent papers. Nevertheless, the author could add the following papers to increase the informational value of this review and pay attention to such issues as:
 - i. these mentioned in Weaknesses section
 - ii. the thorough comparison of various machine learning models in very similar problems of using ML models forecasts in algorithmic investment strategies: Grudniewicz J., Ślepaczuk R., 2023, Application of machine learning in quantitative investment strategies on global stock markets, Research in International Business and Finance, Volume 66, October, <https://doi.org/10.1016/j.ribaf.2023.102052>
- 10. **Data description:**
 - a. all data used in the research should be described in detail in this section, with special attention to:
 - i. frequency, -> **should be added**
 - ii. data range -> **should be added**
 - iii. the rationale for using the specific time series instead of many other possible -> **should be added**
 - iv. The rationale for using the specific assets instead of many other possible -> **should be added**
 - v. descriptive statistics of the data used -> **should be added**
 - vi. a detailed description of data sources
 - vii. A description of any data transformation performed on the raw data set -> **should be added**
 - b. This section almost does not exist in this paper.
- 11. **Methodology:**
 - a. a sufficient and detailed description of the methodology used in the paper and how it was used in order to verify the main research questions and hypotheses. Currently, in this paper, we have only a set of definitions in each chapter and subchapter, mostly without any reference to the content of this paper and details of this research.
 - b. Only one in-sample (IS) and one out-of-sample (OOS) period is not enough to train and test the model on many various periods that characterize different distributions of returns -> then our model is fitted to only one scenario.
 - c. In the case of more complex models (e.g., ML), where the hyperparameter tuning process is a crucial step within the training and estimation model phases, we have to divide our sample into an additional validation period, which is between the training and testing periods.
 - d. No hyperparameter tuning phase is described in this paper, i.e., how empirically was it performed?
 - e. No description of finally selected hyperparameters, at least their more detailed description,
 - f. No explanation of the ranges of hyperparameters tuned or parameters optimized in the training period
 - g. wrong optimization/loss function in the case of models developed for algorithmic strategies. The use of classical forecast error metrics should be replaced by the proper loss function, which takes into account the real returns, which are the consequences of the forecasts from theoretical models, instead of just the comparison between realized and forecasted values. The details of the proper approach can be described in Michańków J., Sakowski P., Ślepaczuk R., 2022, The comparison of LSTM in algorithmic investment strategies on BTC and SP500 index, Sensors 22, 917, <https://doi.org/10.3390/s22030917>
 - h. There are no tests of statistical significance. Without statistical significance tests, it is impossible to determine if the calculated profits or differences in predictive performance are obtained by chance or driven by the tested investment strategies. The same applies to the comparison of forecasts from various models.
 - i. The adequate selection and explanation of the benchmark models
 - j. The lack of transaction costs, especially in the case of research focused on investment strategies, especially in cases where the frequency is higher than daily, can completely distort the results and be the reason for not drawing proper conclusions. Such transaction costs on a real level are especially important if we refer to the practical application of forecasts from ML models in algorithmic trading.
 - k. The use of proper performance metrics to evaluate the efficiency of forecasts from tested theoretical models in algorithmic investment strategies, like for example in: Kość K., Sakowski P., Ślepaczuk R., 2019, Momentum and Contrarian Effects on the Cryptocurrency Market, Physica A: Statistical Mechanics and its Applications 523, pp. 691-701, <https://doi.org/10.1016/j.physa.2019.02.057> ,
 - l. The empirical research should cover a detailed sensitivity analysis (SA) of obtained results to refer to the robustness of tested models concerning the main aim of the research.
- 12. **Results/Conclusions:**
 - a. The main important thing in the case of conclusions is that they should be formulated based on the presented results, proofs, etc., and they should have a rational basis.

- b. The analysis of results should be complete and accurate.
 - c. The reliability of the presented results and validity of withdrawn conclusions should be clearly supported by the data.
 - d. Please check if this section covers the following parts:
 - i. Short description of the problem, main aim, and the whole research.
 - ii. The reference to the main results with a special focus on verified hypotheses and research questions.
 - iii. Policy implications of this research should be added in the conclusion section.
 - iv. Research extensions
13. **General comments:**
- a. All tables in the paper should be self-sufficient. Authors should add a caption for each table that explains its objective and content. The caption should include a brief description of the variables used in the table. Readers should be able to understand the table without going back to the text. Please proceed in the same way for the figures.
 - b. All tables and figures should be referenced in the text of the article.
 - c. The numeration for Figures and Tables is not continuous.
 - d. The numbers of Figures are sometimes wrong, e.g., Figure 8 (in line 760) is supposed to be Figure 13.
 - e. The organization of the manuscript should be rather improved.
14. **Quality of English Language:**
- a. The whole text should be checked with regard to the quality of the English language.
15. **Summary of the review:**
- a. Accept in Present Form
 - b. Accept after Minor Revision
 - c. Reconsider after Major Revision
 - d. Reject