**Title**: Response to Reviewer Comments

**Manuscript Title**: **A Comparative Study: In-Memory vs. Disk-Based Computing with Random Forest for Stock Analysis**

**Date**: 20 April 2025.

We, the authors of this article, would like to sincerely thank the reviewer for their valuable feedback and constructive suggestions. We have carefully revised our manuscript to address all the concerns raised. Below, we provide a point-by-point response to each comment, along with a description of the corresponding changes made to the manuscript.

**Major Comments**

1. **The dataset is insufficiently described - it must be improved. The dataset must be attached to the article as an (online) Appendix.**

Response: We have **expanded the dataset description** in Section [3.1] of the revised manuscript to include details such as the data source, time range, number of records, attributes/features, and pre-processing steps performed before model training. To provide a visual understanding of the dataset, we have included a **representative screenshot** showing a portion of the dataset. Additionally, we have included the complete dataset as an **online Appendix**, as requested.

1. **Where can I find the implementation (source codes of application)? There should be a link or it must be attached to the paper as an online appendix.**

Response: We have provided a **link to the complete source code** of the implementation, which includes data pre-processing, model training using the Random Forest algorithm, and performance evaluation on both Apache Spark and MapReduce frameworks. The source code is hosted on a publicly accessible **GitHub repository**, ensuring easy access and reproducibility of the research.

1. **There are no explicit research questions in the article. It is necessary to add research questions and explicitly answer them in the Results section.**

Response: We have now explicitly formulated the key research questions in the revised manuscript. These questions have been added at the end of the Introduction section to guide the reader and establish a clear focus for the study. Our Results and Discussion section already addressed these research questions in detail.

1. **Contributions to practice and implications for science should be added to the Conclusion section**.

Response: Thank you for this insightful suggestion. We have revised the Conclusion section to explicitly highlight the practical contributions of our research, particularly how the comparative evaluation of in-memory and disk-based distributed computing frameworks can inform real-world applications in financial forecasting and big data processing.

1. **I found a large amount of literature on the subject, which the authors ignore:** [**https://scholar.google.fr/scholar?q=MapReduce+versus+Apache+Spark+in+Stock+Analysis**](https://scholar.google.fr/scholar?q=MapReduce+versus+Apache+Spark+in+Stock+Analysis) **Why don't they discuss this literature in Section 2? Here, they should explain how their research is original compared to these articles.**

Response: We thank the reviewer for pointing out the need to engage more deeply with existing literature. We have **revised Section 2** to incorporate and discuss several key studies from the literature on the comparison of MapReduce and Apache Spark in stock price analysis and big data processing. Furthermore, we have clearly stated the novelty of our research, emphasizing our specific focus on evaluating predictive accuracy and computational efficiency using the Random Forest algorithm within both frameworks, using a real-world stock price dataset.

1. **At the beginning of section 2, a few sentences must describe how the authors selected relevant literature for this section (how the literature review was created).**

Response: Thank you for this helpful suggestion. In response, we have revised the beginning of **Section 2** to include a brief explanation of our **literature review methodology.**

**Minor Comments**

1. **Don't use the link (Ghayoumi, 2023) to references in the abstract.**

Response: Thank you for pointing this out. We have removed the in-text citation (Ghayoumi, 2023) from the Abstract, in accordance with standard academic conventions that discourage the inclusion of references in the abstract section.

1. **References must be in APA 7 style.**

Response: Thank you for bringing this to our attention. We have carefully **reformatted all references** in the manuscript to conform to **APA 7th edition** style guidelines.

1. **Many references are not mentioned in the text! All references must be used in the paper.**

Response: Thank you for highlighting this issue. We have carefully reviewed the reference list and have **tried to incorporate as many of the cited references** as possible into the text. However, in some cases, we found that certain references did not directly align with the content or focus of the paper. If the journal has specific guidelines on the number of references that should be cited or how they should be incorporated, we are happy to make further adjustments.

1. **Provide more context on the specific challenges faced in stock price prediction using big data would strengthen the argument for why this comparison is important.**

Response: Thank you for this valuable suggestion. In response, we have expanded the **Introduction** and **Literature Work** sections to provide a more detailed explanation of the **challenges faced in stock price prediction** when using big data.

1. **Improving the motivations**

Response: Thank you for your feedback on improving the motivations of the study. In response, we have expanded the **Introduction section** to more clearly articulate the significance of our research.

1. **In the introduction part, you give more explanation for MapReduce and Apache Spark.**

Response: Thank you for the suggestion. In response, we have revised the Introduction section to provide a more detailed explanation of MapReduce and Apache Spark. Specifically, we have added: (a) An overview of MapReduce, explaining its design principles, how it processes large-scale data in a distributed manner, and its applications in big data analytics. (b) A more detailed description of Apache Spark, including its in-memory processing capabilities, advantages over MapReduce in terms of speed and flexibility, and its popularity in machine learning and data processing tasks.

1. **For review part, you should improve the link between studies and your study objectives and more clarified.**

Response: Thank you for this suggestion. In response, we have revised the **Literature Review** section to better link the **existing studies** with our **study objectives.**

1. **You should describe more details for dataset.**

Response: Thank you for this helpful suggestion. In response, we have expanded the **Dataset** description in the revised manuscript.

1. **Adding more specific details on the evaluation metrics to strengthen the explanation.**

Response: Thank you for the insightful suggestion. In particular, we have:

* Clearly defined each evaluation metric used, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²).
* Explained the purpose and interpretation of each metric in the context of stock price prediction.

1. **In the discussion section, you should provide more details and compare your findings with other studies.**

Response: In response, we have revised the Results and **Discussion** section.

1. **In the conclusion, you should discuss more about the applications of the research.**

Response: In response, we have enhanced the **Conclusion** section by elaborating on the **practical applications** of our research. Additionally, we have included a **Future Work** subsection that outlines potential directions for extending our research.

We sincerely thank the reviewer for their detailed and constructive feedback. We have tried our best to carefully address each of the points raised and made the necessary revisions to improve the quality and clarity of our manuscript. The changes have been incorporated as per the suggestions, and we have provided a point-by-point response above.

We hope that the revised manuscript now meets the expectations for publication. If any further changes are required, we would be happy to revise the paper accordingly.

Thank you for your time and consideration.

Sincerely,  
Chitra Joshi.

On behalf of all co-authors