Response Letter for Adding Authors

Dear Editor:

First of all, I would like to express my sincere gratitude to you and the review team for your professionalism and patience during the review process. We are aware that it is a great honour to have the opportunity to publish our research in Peerj Computer Science, and we sincerely look forward to it.

Regarding your question about the addition of authors during the submission process, I would like to provide you with the details and rationale. As this study enters a critical phase, I realise that we need to strengthen our expertise in computing and finance to go deeper, and at the same time we need to conduct more extensive experimental verification to ensure the accuracy and depth of our research. So we decided to invite Professor Tang Yong and Mr Huang Jiacheng to join my research team.

Professor Tang Yong is the deputy director of the Institute of Economic Research at Fuzhou University, has published more than 30 academic papers on financial engineering and Internet finance, and has led a number of scientific research projects, including the National Natural Science Fund Project and the Ministry of Education's Humanities and Social Sciences Project. He has provided valuable guidance and insight to our research, especially in the technical details of natural language processing, which has greatly improved the quality and depth of our research. Under Professor Tang's guidance, we have explored natural language technology in greater depth and provided valuable constructive suggestions for the future work of this research with a fresh perspective to enhance the usefulness of LEET. Meanwhile, the National Social Science Foundation of China has provided financial support for this research under grant no.21BJY033, chaired by Prof Tang.

What follows is a relevant introduction of the contributions made by Prof Tang Yong:

1. More detailed elaboration and implementation of the techniques of natural language processing to enhance the experimental expertise and reproducibility, as follows:

***Data preprocessing***

Since news headlines may contain superfluous information, such as punctuation marks, special characters, repeated words, or suspended words, this experiment aims to eliminate these elements and refine the data of news headlines (Usmani&Shamsi, 2023). Meanwhile, the author performs lexical segmentation and annotation, removes inactive words and reduces them to their original or stemmed forms, to standardise news headlines and improve the accuracy of natural language processing.

Meanwhile, we use the roberta.large model (Liu et al., 2019), a pre-trained deep learning model for natural language processing, to implement sentiment classification. First, we use roberta.large model's tokeniser to segment text into tokens; the tokeniser tries to split words or text fragments into smaller units until all units can be found in the vocabulary. Then, to identify the start and end of a sentence sequence, roberta.large model uses <s> as the start token of the sequence and </s> as the end token of the sequence.After splitting into tokens, these tokens are converted into indexes by looking up the corresponding unique index of each token in roberta.large model's vocabulary. Since the roberta.large model requires that all input sequences have the same length, shorter sequences must be padded and longer sequences must be truncated. In order for the model to know which positions are real tokens and which positions are padded tokens, an attention mask must also be created. For real tokens, the mask value is 1; for padded tokens, the mask value is 0. The above steps convert the raw text into a format that can be processed by the roberta.large model.

2. To explore the future of this research in terms of helping investors to make rational decisions, as described below:

**FUTURE WORK**

**Model performance**

To help investors make more intuitive decisions in the future, we plan to expand our research and development efforts in several key areas:

1. Integration of additional economic indicators

While sentiment analysis and sentiment indices provide valuable insights into market trends, the integration of additional economic indicators can significantly improve the accuracy of a model. These indicators can include macroeconomic data such as GDP growth rates, inflation rates. By analysing the interplay between market sentiment and these economic fundamentals, our models can provide a more complete view of the market and thus predict stock returns with greater accuracy.

2. Extension to global equity markets

Given the interconnected nature of today's global economy, events in one region can have a profound impact on global financial markets. We therefore plan to extend our model to global equity markets, including emerging markets.

3. Real-time data processing

To capture the dynamic nature of financial markets, our future research will focus on implementing real-time data processing. This will allow models to incorporate the latest market news, social media trends and economic reports as they become available. Real-time processing will make our models more sensitive to sudden market changes, providing investors with timely insights to guide their trading decisions.

4. Advanced natural language processing technology

We plan to explore the use of more advanced natural language processing (NLP) technologies that go beyond sentiment analysis. These technologies could include context-aware models capable of understanding the nuances of financial news, such as the difference between the short-term and long-term impact of an event.

5. User-Friendly dashboard

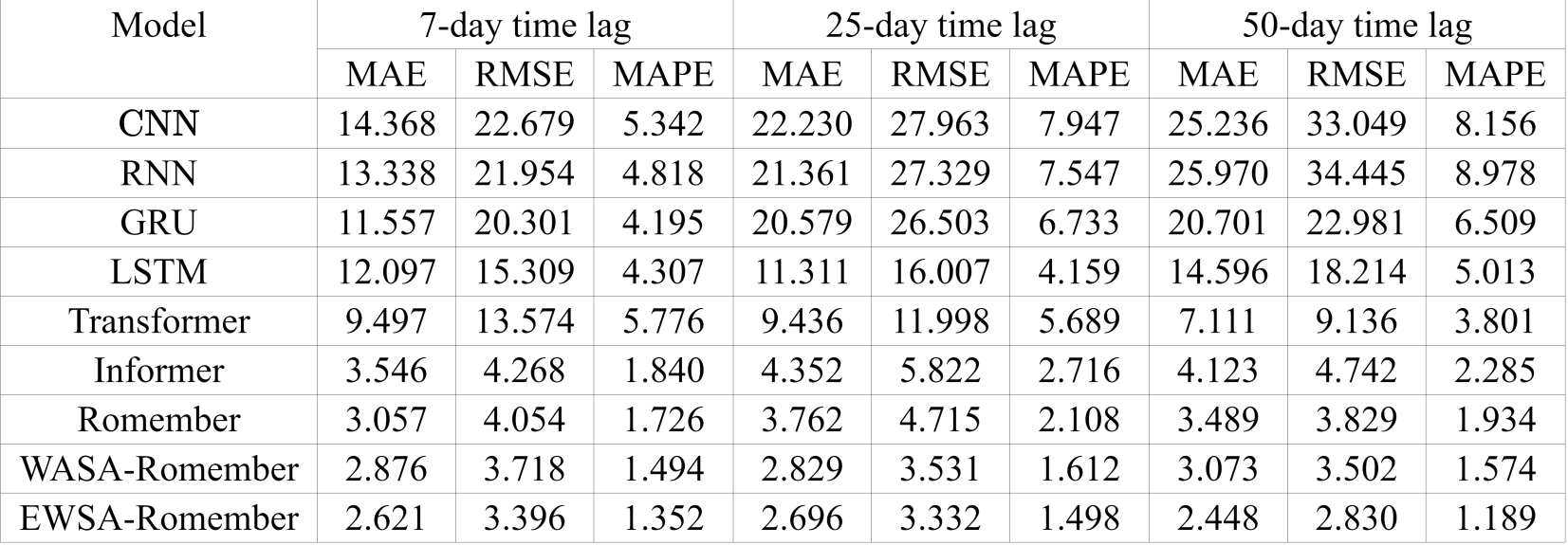
We plan to design an intuitive dashboard that presents a holistic view of the market sentiment and trends, integrating the LEET method's insights with additional economic indicators. This dashboard will feature easy-to-understand visualizations, such as sentiment trend graphs, economic indicators charts, and predictive analytics insights. It will enable users to quickly grasp the current market sentiment, understand how it correlates with key economic indicators, and foresee potential market movements.

Mr Huang Jiacheng, a member of the Menus International School of Engineering at Fuzhou University, has particular expertise in experimental research and innovation, while his contributions to experimental design and data analysis have been crucial in ensuring the accuracy and reliability of our findings. He extended our experimental thinking and conducted our experiments using the new FTSE100 dataset, which is representative of the UK market, to refine our experiments. He also provided insight into the ethical aspects of the study. The addition of Mr Wong to our team has not only enhanced our technical capabilities, but has also brought new perspectives and ideas to our research work, and significantly increased the social responsibility of our research.

Next is a description of Mr Wong Ka Shing's contribution.

1.The FTSE 100 was used as a supplementary dataset for additional experimentation and validation of the LEET methodology in this study:

**Table 2. Indicators with 5,25,50 day time lag based on FTSE 100**

****

2. To explore and discuss future work on the ethical aspects of this study:

**Ethical aspect**

While the LEET method shows promising results in using emotional analysis for stock market predictions, it is imperative to discuss the ethical and legal implications associated with such methodologies, particularly with respect to privacy and market manipulation.

First of all, the use of emotional analysis in financial markets raises significant privacy concerns. The EWSA and WASA indices, which are fundamental to the LEET methodology, rely on vast amounts of data extracted from various sources, including social media, news outlets and forums. This data collection process must comply with strict data protection laws, such as the General Data Protection Regulation (GDPR) in the European Union and the California Consumer Privacy Act (CCPA) in the United States. It is crucial to ensure that data is collected and processed transparently, with the explicit consent of individuals, to avoid violating their privacy rights.

In addition, the application of the Romember long order sentiment time series model within the LEET framework requires ethical consideration of how emotional analysis could potentially be exploited for market manipulation. The predictive capabilities of the LEET method, while beneficial for understanding long-term market sentiment, may inadvertently create opportunities for information asymmetry. This asymmetry may enable certain investors to influence market trends based on sentiment analysis, thereby undermining the integrity of financial markets and violating principles of fair trading.

To address the ethical and legal challenges posed by sentiment analysis in stock market forecasting, particularly in terms of data privacy and market manipulation, this study provides a relevant discussion and suggests future improvements:

**Transparent use of data**

**1.Data source disclosure:** When using sentiment analysis tools such as the LEET method, clearly list all data sources, including social media platforms, news sites, forums, etc., and explain how this data is collected and processed.

**2.User consent for access:** Ensure that explicit user consent is obtained before personal data is collected. This can be done by updating privacy policies, providing easy-to-understand consent forms, etc.

**Ethical guidelines for sentiment analysis**

**1.Establish ethical guidelines:** Develop a set of ethical guidelines specific to the use of sentiment analysis in stock market forecasting, including but not limited to protecting individual privacy, ensuring data accuracy and avoiding bias.

**2.Avoid market manipulation:** Explicitly prohibit the use of sentiment analysis results for market manipulation or unfair trading practices. This includes preventing trading on the basis of undisclosed information or using sentiment analysis results to mislead other investors.

**Regulatory compliance and oversight**

**1.Working with regulators:** Proactively communicating with financial market regulators to ensure that sentiment analysis methods comply with existing market rules and regulations.

**2.Regular audits:** Conduct regular internal or third party audits to ensure that sentiment analysis tools are used in accordance with ethical and legal requirements, particularly with regard to data protection and market fairness.

**Investor education**

**1.Transparent risk disclosure:** Clearly disclose the potential risks and uncertainties when investors use forecasting tools based on sentiment analysis.

**2.Promoting rational investment:** Encouraging investors to make comprehensive and rational investment decisions by combining the results of sentiment analysis with other investment information such as fundamental analysis.

We understand your concerns about the change in the list of authors on the manuscript and have ensured that this decision was taken after careful consideration and discussion with the aim of improving the quality and integrity of our research. We have ensured that all authors have reviewed the final manuscript and agree to take responsibility for its content.

We hope that the above explanations have allayed your concerns, and we stand ready to provide any further information or clarification to facilitate your and the review team's final assessment of our manuscript.

Once again, we thank you for your understanding and support. We look forward to hearing from you and hope to continue the process of publishing our findings in your journal.

Yours sincerely，

Honglin Liao

onglinguge@gmail.com