**Peer review sheet**

MAFS6010Z, 2021 fall

Your name and sid: LU Qiaoyu (20736916)

Group that you review: 9

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Confidence on your assessment (1-3) | Clarity and quality of writing (1-5) | Technical quality  (1-5) | Overall rating  (1-5) |
| Score | 3 | 4 | 3 | 4 |

Summary:

The report of Group 9 introduced their replication method towards the original paper. They described the data and visualized the distribution of some variables used. The report also explained how they analyzed and handling the missing values clearly. In the Methodology part of the report, initially they included their dataset splitting and performance evaluation method. Besides, they presented several methods used to fit the models, such as Simple Linear (OLS and OLS with Huber), PLS, PCA, Random Forest, GBRT and Neural Network, as well as the underlying logics. Next, they analyzed their model scores and the feature importance. Finally, in view of the model limitations, they gave some suggestions for improvement.

Strengths:

The report introduced their modeling process with a complete structure. They provided profound and suggestive reflections on the shortcomings of models. Besides, they displayed the variable correlations and detected the inner connection between several variables. They also paid more attention to missing value handling and standardization such that well preprocessing the data. For their code part, the functions designed to fit models and compare the performance are concise and clear, which are easy to understand.

Weaknesses:

In my point of view, they could include some models presented in the original paper, such as Ols-3(Huber) and Elastic Net with Huber, which may be more closed to the research. Besides, they could adopt parameter adjusting part into their recursive evaluation scheme. I noticed that some parameters like L1-ratio and L2-ratio in Elastic Net are fixed, which may influence the test scores. Additionally, they used R-Square score in ‘sklearn’. But the definition of R-Square score may be different from that in the original paper.

Clarity and writing:

The report was clearly written and well-organized figures. But they could present more materials for further analysis on specific models. They could display variable importance analysis on each model.

Technical quality:

The technical quality of this report is good, but as mentioned in ‘Weaknesses’, they could replicate more models in original paper to improve the scores.