**Peer review sheet**

MAFS6010Z, 2021 fall

Your name and sid: Wong Sik Tsun (20038819)

Group that you review: Group 19

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

Summary:

The report summarized how the group replicated the paper to use Lasso, Ridge, SGD, PCR, GBRT and random forest methods to predict stock market data through recursive evaluation and use R squared to evaluate the effectiveness of each method and the importance of each variable. The report is technically sound and clearly written. Nevertheless, there are also several areas that improvement can be made, for example the handling of missing data and the clarity of the code.

Strengths:

1. The group replicated the 6 models using similar approach as suggested in the paper for the recursive performance evaluation scheme, hyperparameters tuning, R squared calculation and feature importance calculation. The result is clearly presented.
2. Variable selection was performed based on the correlation analysis and K-means clustering analysis, which helps reducing the data dimension.
3. The codes are well structured.

Weaknesses:

1. The group handled the missing data by replacing them with 0 for firm characteristics and removing samples with missing returns. Such approach is different than that in the paper, which replace the missing characteristic with the cross-section median.
2. The robust loss function is not mentioned in the report.
3. The group did not consider the interaction among features, but it is understandable given that the number of samples and features is very large, which may cause memory issue when the interaction terms are considered.

Clarity and writing:

1. The report is clearly written. It is organized which mentioned the data processing first, followed by the introduction of the models, and then the empirical analysis. It also includes tables and figures to help illustrating the content.
2. The codes are well structured, but it is suggested to add comments so that the reader can understand the purpose of the codes more easily.

Technical quality:

1. The result is technically sound and have no obvious flaws in the reasoning. The methodology is clearly presented in the report. However, more work can be done in the data preparation part, for example the missing data handling.