**Peer review sheet**

MAFS6010Z, 2021 fall

Your name and sid: HE Haokai, 20744858

Group that you review: 14

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Confidence on your assessment (1-3) | Clarity and quality of writing (1-5) | Technical quality  (1-5) | Overall rating  (1-5) |
| Score | 3 | 3.5 | 3.5 | 3.5 |

Summary:

Group 14 replicates part of the study of Empirical Asset Pricing via Machine Learning. Unlike the original study, group 14 changes the method of filling missing values and remarkably decreases the number of interaction variables. Group 14 finds that GBRT, RF and NN5 has the highest R2 score among models of PLS, PCR, Elastic Net, RF, GBRT, NN1, NN2, NN3, NN4 and NN%.

Strengths:

Innovatively applying feature selection to greatly reduce the number of interaction variables by selecting high correlation features, which effectively resolve the problem of huge memory footprint during model estimation in the original study.

Weaknesses:

The report is too short to introduce more details of data processing and parameter selection of various models.

Clarity and writing:

There are several fundamental grammatical and spelling mistakes in the report. Besides, the report has an inaccurate statement that interaction variables play well in many models means these variables are significant.

Technical quality:

The study of group 14 doesn’t cite the source of data of 8 macro variables. Details of basis for selection of the parameters of Neutral Network models are missing both in report and in source code. Discrepancy of variable importance across different models is unclear since the report only displays individual graphs of variable importance of each model.