Peer review sheet

MAFS6010Z, 2021 fall

Your name and sid: Li Xintong (20736227)
Group that you review: Group 14

	Confidence on your assessment (1-3)	Clarity and quality of writing (1-5)	Technical quality (1-5)	Overall rating (1-5)
Score	2	4.4	4.3	4.3

Summary:
Group 14 chose 6 models from the paper: Elastic Net, PCR, PLS, RF, GBRT and NN. They eliminated variables which do not appear in the reference and re-coded sci2 by one-hot code. They dropped data without price and the rest of the missing values were filled with 0. Features were selected by Pearson correlation coefficient and by multiplying 30 selected stock features and 8 macroeconomic features, they finally used total 270 features to train the models. For each model, they calculated the reduction in R^2 from dropping a given predictor to get characteristic importance and divided the top-20 most influential variables four categories. For model performance, GBRT, RF and NN5 are top 3 models according to their result.

Strengths:
Grouped the top-20 most influential variables into 4 categories, which gives more insights about the reason why these features are important. The results of some models like NN5, GBRT, RF are closed to the paper.

Weaknesses:
Didn't calculate the out-of-sample R^2 for top market-value stocks and the bottom ones which is required in the ppt that teaching assistant gave. Feature selection process has space to be improved. Maybe could do more than just calculating Pearson correlation coefficient.

Clarity and writing:
The poster is good over all in format and the steps in the projects are clearly addressed in different blocks. But there are some typos like "anf" in part 6.

Technical quality:
Technical quality is great over all. Used quantitative index to select variables and the code is nicely commented.