# MAFS6010Z: Project1 Review for Group 4

Yu Xintong  xyubi@connect.ust.hk(20745503)

**Summary of the report:**

The group went through the variable selection, model selection and model implementation procedures. In variable selection part, they chose variables by experience and correlation analysis and then standardized them. In model selection and model implementation procedures, they evaluated Logistics Regression, KNN, Random Forest and Gradient Boosting models, and tested which one performed better on Kaggle. Finally, they chose random forest.

**The strengths of the report:**

1. Careful selection of variables: The report shows that they investigated the variables carefully to see if the variables have Economic-related impacts. This is a job that requires experience and patience
2. More model options: They chose four models to compare and evaluate which works better, which can help to select a more suitable model.

**The weaknesses of the report:**

1. A more scientific approach to variable selection could be used: Using only experience for variable selection may cause us to lose important variables. They can calculate the contribution of variables to the target value and continuously adjust the model
2. Need to be careful when use test data for model selection: The report finally chose Random Forest for it performed better on Kaggle. But using test data for model selection may be equivalent to converting test data into training data and requires care.

**Evaluation on Clarity and quality of writing:**

Score:4.8 or 5(int)     The report is clear and well organized

**Evaluation on Technical Quality**

Score:4.5 or 4(int)
The overall technical quality of the report is very good, and there are some methods that could be improved in the variable and model selection process.

**Overall rating:** Score: 4
**Confidence on your assessment:** Score: 3