# Evaluation on Project 15 (MAFS6010Z)

**Student Number: 20568852**

## Overall: 4 points
## Confidence: 3 points

### Summary of the report

The group picked 4 more datasets besides the main table, namely "bureau", "POS_CASH_balance", "previous_application" and "credit_card_balance" for feature selection and feature engineering processes. They use those data to predict whether an applicant will be able to repay a loan.

### Describe the strengths of the report

The report uses more features from other datasets and try PCA to do feature selection. It also uses the logistic regression model as a baseline to testify the accuracy of random forest and lightGBM model.

### Describe the weaknesses of the report.

While in the training data and in the testing data, different classes are distributed with huge bias, where the percentage being around 10:1 and that will cause the accuracy on the testing data seems good. Some methods to deal with the unbalance should be implemented.

### Evaluation on Clarity and quality of writing (1-5):

Overall: 5 points

Confidence: 3 points

- Is the report clearly written? Yes
- Is there a good use of examples and figures? Yes
- Is it well organized? Yes
- Are there problems with style and grammar? No
- Are there issues with typos, formatting, references, etc.? No

### Evaluation on Technical Quality (1-5):

Overall: 5 points

Confidence: 3 points

- Are the results technically sound? Yes
- Are there obvious flaws in the reasoning? No
- Are claims well-supported by theoretical analysis or experimental results? Yes
- Are the experiments well thought out and convincing? Yes
- Will it be possible for other researchers to replicate these results? Yes

- Is the evaluation appropriate? Yes

- Did the authors clearly assess both the strengths and weaknesses of their approach? No

- Are relevant papers cited, discussed, and compared to the presented work? No