

## **MAFS6010Z Project 1 Peer Review**

Reviewer: LU Qiaoyu(Group 19)

Group to review: Group 1

### **Part 1: Summary of the report**

Initially, the report of Group 1 described data preprocessing and feature engineering methods in detail, including handling outlier and missing values, label and one-hot encoder, constructing polynomial features, dimensionality reduction and normalization. Then the report mainly introduced corresponding principles of several models selected to use in the Home Credit Default Risk competition, such as Logistic Regression, Random Forest and LightGBM. Finally, the report compared respective results of different models and concluded that.

### **Part 2: Strengths of the report**

Firstly, the report introduced the process of data selection, as well as principles of feature engineering methods and models used, clearly indicating the underlying logic and reasons of the project. It also referenced some formulas for easily understanding. Secondly, the last part of the report included the process of how Group 1 tried to improve scores, implying the authors' clear and scientific mind. Thirdly, some figures and tables properly illustrated the process of data analysis, which were used as powerful evidence of experimental claims.

### **Part 3: Weaknesses of the report**

From my personal point of view, the report could give a discussion on the model selection. Rather than just showing out the results, it could provide an evaluation of strengths and drawbacks of each model or why various models lead to different results. Based on the analysis, it may reflect and propose improvement measures. In addition, it could provide any evidence to analyze how feature engineering improve the model efficiency.

### **Part 4: Evaluation on Clarity and quality of writing: 5**

The report is clearly written and well organized.

### **Part 5: Evaluation on Technical Quality: 4**

The report gave scientific explanation for all methods used. However, the authors did not include a discussion on assessing strengths and weaknesses of their approach.

### **Part 6: Overall rating: 4**

Above all, my vote is 4-a good report.

### **Part 7: Confidence on my assessment: 3**

I have carefully read the paper and checked the results.