

Group 7

1. Summary of the report

Shao, Zhihao is working on the Home Default Risk dataset, and he focuses on the question about whether the model should be fed with the full set of features without selection or a small subset after close examination. He first wrote a brief introduction, then he gave a brief description of the dataset. He then gave some exploratory analysis on the data and he conducted the PCA analysis and chose the top 40 components. At last, 16 predictors with highest correlations were selected. He then mentioned which models he has used, which are logistic regression, linear discriminant analysis, and random forest, and some AUC graphs to visualize his results. He then gave the analysis, conclusion, and his reference at last.

2. Strengths of the report

- (1) He put lots of graphs in his report, for example, the pie chart for target distribution, correlation heatmap, AUC curve for the result, etc., which is good for visualization
- (2) He gave quite an insightful analysis
- (3) He did provide the reference he used for the report

3. Weaknesses of the report

- (1) He did not mention a lot about the models, and the reasons why he chose those models, and the reasons why he set the model parameters to be those numbers
- (2) The graphs are small so that it is hard to see the numbers and words written in the graphs
- (3) He only explicitly gave the best result, but he didn't specify the result for other models, also, he didn't mention which model has the best result

4. Evaluation on quality of writing (4)

I think the report is clearly written as it did express his thoughts clearly, and there were figures for visualization and there were no typos. I think the parts that this report can improve are as follows :

- (1) For the second point of the analysis, the first letter should be capitalized as other points have capital letters as their first letters.
- (2) He could explain more for exploratory data analysis, as he mentioned he analyzed the data's normality, multicollinearity, missing value status but he didn't mention how he performed his analysis.

5. Evaluation on presentation (3)

The presentation was clear and well organized, but the language flow was not very fluent. The slides were clear and well prepared as he used different colors for highlight points, and he put lots of graphs in the presentation which brings a great

visualization. Also, he gave a lot of analysis at last. However, there are some points that I think he can improve, which are as follows :

- (1) The wordings in the graph that he put in the presentation were too small, and I can barely see the words in the graph
- (2) He was quite nervous when presenting

6. Evaluation on creativity (2)

I don't think the work proposes any new ideas, as I believe these methods have been proposed for the Home Credit dataset before, hence, I don't think that this report contains any state-of-the-art results. Also, the accuracy for the best result was not very impressive, hence, I would not utilize the parameters of his random forest model.

7. Confidence on my assessment (3)

I have carefully read the paper and checked the results.