Group 4

• Summary of the report.

Group 4's report shows two approaches to the Home Credit's problem of predicting their clients' repayment abilities based on alternative or historical data. Specifically, they use Logistics Regression and light Gradient Boosting Machine to make predictions and provide some analyses on the results. Their report begins by briefly introducing about the dataset, how they pre-processed it and which features they pick into their models. Subsequently, they vaguely describe about their selected models together with their Kaggle results. Finally, they provide some corresponding observations about their approaches and make some reasonable conclusion.

• Describe the strengths of the report.

The strength of Group 4's report lies on the Conclusion part where they make some meaningful insights about the reasons why the light GBM model performs better than the Logistics Regression model, and also the implication that data comprehension can lead to better prediction.

• Describe the weaknesses of the report.

There are several issues about the report that I want to point out:

- Lack of inference that relates to domain knowledge: I think we would need to look at all attribute of the data set, then approximate significance of each feature base on domain knowledge. We cannot just rely on the ML model.
- Not provide the structure of data files as in presentation, not describe clear about merging the test data set, and how to merge it
- In the second statement of Feature Selection, correlation is not enough to judge the relationship of a feature and the target (many cases can happen, as we face them regularly in Statistics) -> may need some pairwise scatter plot.
- May work more on model, like try other models, tune the parameters, ... Applying just two light models and comparing them may be a bit simple

• Evaluation on quality of writing (1-5): Is the report clearly written? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper, and provide details of typos. 4

Evaluation is as follow and Group 4 can improve based on this.

- The report is somewhat clearly written
- They did not really make good use of illustration. The figures are a bit lost in the poster with no description of what purpose they are used for. And the table of results is incorrectly matched with what they mention in the section of dataset, in which they said they utilized only the application dataset for Logistics Regression.
- It is quite well-organized
- No problems with style and grammar
- They made small typo in marking their sections' title and their reference seems to be inappropriate cited

• Evaluation on presentation (1-5): Is the presentation clear and well organized? Are the language flow fluent and persuasive? Are the slides clear and well elaborated? Please make suggestions to improve the presentation. 5

Evaluation is as follow and Group 4 can improve based on this

- The presentation is clear, well organized, and aligned with the report
- They present fluently with clear slides

• Evaluation on creativity (1-5): Does the work propose any genuinely new ideas? Is this a work that you are eager to read and cite? Does it contain some state-of-the-art results? As a reviewer you should try to assess whether the ideas are truly new and creative. Novel combinations, adaptations or extensions of existing ideas are also valuable. 3

Evaluation is as follow

- Their work doesn't propose any novel idea
- There is a spark about their purpose of using Light GBM but they only applied it in a very simple way with no model tunning.
- We did expect that they use more models and then compare their performances.

• Confidence on your assessment (1-3) (3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong) 3

I have carefully read the paper and checked the results (but not run the model again)