MSBD5013 Midterm Project

Peer Review for Group 10

1. Summary of the report

Group 10 makes full use of data analysis, feature engineering techniques which also provided good supporting ideas to read. For example, they detected the feature distributions and drop the missing value column features in a reasonable way. They also add several features for enrichment. With multiple models involved, L1 regularization and stratified 5-fold cross validation, they compared and searched ultimate 4 models for prediction. Finally, a good interpretation is also discussed in this report for feature selection analysis.

2. Strengths of the report

(1) The report is of good logistic and well-organized.

(2) The workflow of this report can be easy to follow-up.

(3) This report involves rich support ideas and reasons for feature engineering and model selection, which makes full sense.

(4) Multiple models such as KNN, LightGBM, Random Forest, XGBoost are involved for performance comparison.

(5) Good interpretation using SHAP method of LightGBM feature selection analysis.

3. Weakness of the report

(1) More details about aim and the datasets for mainly usage can be involved in introduction part.

(2) In experimental results part, if you rank the performance of the model and highlight the most accurate and efficient model, I believe it will be also good for model interpretation.

(3) A friendly reminder that a litter typo in contribution part about the second group member.

4. Rating of the report

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

4

This is a well-organized report and there are also no problems with style, grammar, basically. Besides, detailed reasons and supporting ideas are involved in this report which convince me a lot. A small typo in contribution part, but overall is good enough for me to read. More figures can be illustrated in this report for direct observation in exploratory data analysis.

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

5

The AUC result is sound in prediction. The model can also be easy to replicate. And they also use SHAP method of LighGBM model for feature selection analysis which is good for model interpretation.

(3) Overall rating (1-5):

4.5

It is a good report for me to gain some knowledge.

(4) Confidence on your assessment (1-3):

3

I have carefully read the paper and checked the results.