Peer Review of group 6

Summary of the report

It is a simple report mainly describing how to select and generate features from the original data and what the result is by using LGBM model. The selected features have high correlation with the TARGET, and the threshold is set to be 0.02. However, some information and details are not presented in the report but in the jupyter notebook instead.

Strengths of the report

Simplicity and comparatively high score among classmates.

Weaknesses of the report

Details of the project is not shown in the report. Putting something mentioned in the jupyter notebook to the report is better and adding information in model selection is also a good choice.

Explanations of the steps are not enough, like why to use these selected features.

Evaluation on Clarity and quality of writing: 2

Plots are given to further explain the features, but it seems to be a shallow explanation. Going further to talk about the result will be better.

The report seems not to follow the guidance given on the website. For example, the font is not Times New Roman, and the plots do not have titles.

References are not shown in the last of the report, and some codes in the jupyter notebook are very similar to those given online.

Evaluation on Technical Quality: 2

The selected features have high correlation with the TARGET, and the threshold is set to be 0.02 according to the report. I think it is a good approach.

It is appropriate to use cross-validation method for assessment.

Used methods are stated in the report, but detailed description is missing, such as the principles and advantages of the model.

Some approaches stated in the report seem to have low impact in the experiment, but reasons for these failures are not given, and these parts are deleted in the codes.

LabelEncoder is used in the experiment instead of one-hot encoder, which may lead to some comparison in one feature. For example, those sub-features with larger given numbers may have greater influence.

RandomForest is used in the experiment, but the comparison between LGBM and it misses.

We are able to follow the codes and replicate the result.

Overall rating: 2

Confidence on assessment: 2