Review of Group 4

Summary: The group 4 chose the regression problem. They did a very detailed EDA on the data, analyzing and visualizing the data from different dimensions. They selected LGB for modeling. The result of the model was excellent.

Strengths:

(a) Very specific EDA process involving visualization of different features and their combination.

(b) The explanation of the principle of LGB is very detailed, from the principle of GBDT(including pseudocode) to the explanation of LGB is better than GBDT is very clear.

(c) The model performs excellently, with an error of 0.6 for the private dataset and 0 for the public dataset.

(d) The structure of the report is complete, the description is clear, and there are no obvious errors.

Weaknesses:

(a) Just the LGB. More models can enrich the technical contents and have a contrast with LGB.

(b) Due to the lack of feature engineering (feature selection and feature aggregation--mentioned in report part5), in fact, the EDA is not very helpful for the subsequent modeling process.

(c) Lack of explanation and support for the final result. For example how to complete the parameter tuning to achieve the current performance, the several attempts mentioned in the part3.2 can be expanded.

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

The writing is of high quality, the only mistake I found was the meeting information at the bottom of the front page was not revised.

Evaluation on Technical Quality (1-5):3.5

As mentioned above,the EDA job is quite good, while the other technical contents is not enough, lacking some feature engineering and having just one model of LGB. There is a lack of support for the final result, but the performance is a plus.

Overall rating:4

Confidence on your assessment:3