Review for the report of group 1

Summary

This group takes almost all the given tables into account. In the part of data selection, they drop some outliers and some variables which may be of little importance to TARGRT. In addition, they mainly use median value to fill the missing data. In the part of feature engineering, they use the method of one-hot encoding to encode the categorical features. They also add some polynomial features based on the feature which is highly correlated with TARGET and then do the normalization to the data. After dealing with all the data, they do the model selection. Several models like logistic regression, random forest, light GBM as well as the combined model are taken into account and we can see that light GBM performs best.

Strength

They engage enough features in this project is enough and consider many methods to deal with the data so that the data will show higher efficiency in the future training.

Weakness

The method used to fill the missing value is too simple which may reduce the final AUC. In addition, Adjusting parameters can be engaged in model selection.

Evaluation on Clarity and quality of writing: 4

Maybe the results of more descriptive statistics in the format of table and figures should be shown to illustrate the correlation between some selected variables and TARGET.

Evaluation on Technical Quality: 4

Overall rating: 4

Confidence on your assessment:3