

## Peer Review of Group7

The author did a good job in describing what he did and why. He did a data cleaning work first, filled the low percent missing value with the most common value in that column. Then he used LR and RF to model these data and compared their performance.

The report has a good struture which is very clear and easy to understand, but there are some technical details worth optimizing. In data processing, the author dropped the columns which have high percentage of missing value. In my opnion, there is a possibility that these missing value represent some important information. For example, a guy may have a higher possibility than those who have the house information. And in terms of model, the authur didn't give the evidence that some features have linear correlation with TARGET. So I think it is not a good idea to choose LR as the model, because the number of features is large and the percentage of positive sample is low. RF is better.

## Rating

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

Evaluation on Technical Quality (1-5): 4

Overall rating: 4

Confidence on your assessment (1-3): 3

*reviewed by SixianHUO*