Peer review for Project of Group 2

Name: ZHEN Mengnan

Summary:

The project has some contents about exploring the data relationship using single feature correlation. For the further analysis, qualitative and quantitative variables were divided. The qualitative variables were numerically coded. VIF was used to do the variable election, the team deleted the variables whose VIF is greater than 10.

Then, the team tried different classified methods to fit the data, and they chose LGBM because of higher AUC value. The result of Kaggle test is 0.748.

Strengths:

- Try to fill some missing values
- Try different types of classifiers
- Using RobustScaler to reduce the effect of outliers

Weaknesses:

- Just tried four types of classifiers using packages directly (LDA/LR/GB/SGD in the code), Not very innovative methods.
- Deleted the variables whose VIF is greater than 10 directly, rather than using other methods to decrease the effect of multicollinearity (such as Ridge regression/PCA and some other methods)

i.   Evaluation on Clarity and quality of writing:4
ii. Evaluation on Technical Quality:4
iii. **Overall rating:4**
iv. Confidence on assessment:3