

Peer Review Group 10

1. Summary
 - a) Group 10 did a lot of data preprocessing and feature engineering such as feature correlation analysis, feature integration. Based on the feature analysis, features with 60% missing values, binary features with 90% domination and features with zero coefficient are removed.
 - b) About the models, they focused on tree-based model to solve such as XGBoost, LightGBM and Random Forest. Besides, L1 regularization and k-fold cross validation is used to prevent overfitting.
2. Strength and Weakness
 - a) Strength
 - i. Wonderful feature engineering
 - ii. Discuss the issue of imbalanced label
 - iii. Compare several models' performances
 - b) Weakness
 - i. Too less details about the model, and do not show the details of the hyper-parameters selection
 - ii. More figures and examples will be better
 - iii. Though discuss the issue of imbalance label, they did not solve the problem completely
3. Clarity and quality of writing: 4
The report is clearly written with examples and figures. It is well organized and readily comprehensible without typo and grammatical mistakes. One area that could be improved is adding more figures and experiment results.
4. Technical quality: 4
 - a) The feature engineering as well as the model selection reasoning are both sound and reasonable. But they focus too much on the feature engineering part, which contains more than two third, though is thorough and detailed.
 - b) Several evaluation metrics are used instead of using only ROC scores.
 - c) Although they discussed the issue of imbalanced label, they did not take actions to solve it thoroughly. One possible improvement might be oversampling and desampling.
 - d) The report is kind of concise, maybe better with more technical description in details.
5. Overall rating: 4
6. Confidence: 2