Rebuttal from group 4

Review 5 :
Rebuttal on the weakness of the report :
Our initial thought for the prediction model was logistic regression, and we have used it in our project as the base model, which we have mentioned in our poster.

Rebuttal on creativity:
We do not agree that there is a lack of specificity towards the problem, as the data preprocessing, feature selection, and model selection were in fact specified to this problem. We also spend great effort in the dataset aggregation step, which provides an insight on how to utilize other datasets, and examine the efficiency of including more features in a machine learning model.
We proposed a novel problem statement about analyzing the difference between merged dataset and application dataset, the focus of this project is to examine whether more features can give higher prediction results. Therefore, we adopted two major models to test this hypothesis. And Light GBM is a state-of-art model which was just developed in recent years. It has a high model complexity and accuracy, and it is the most winning model in Kaggle contests. Moreover, we did spend great effort on understanding the deep architecture instead of just using it and we have explained in class, as well as understanding the relationship between predictors and response variables, as shown in P.9-P.14 of our presentation slides. So we think that these two major models are sufficient to draw our conclusion that more features are indeed useful for prediction. Moreover, we did tune our LightGBM model, with fine tuning parameters and grid search for hyperparameters, as presented and explained in P.20 of our presentation slides. Detailed implementation can be referred to our code.
We think this should be considered as adaptations and extensions of existing ideas and counted as creativity.