Group 4 used Home Credit default risk dataset to predict ability of people to repay loans given historical data. They ended up using logistic regression and a gradient boosting model (Light GBM) on the merged dataset and achieved a score of 0.757.

Their report was concise, well-organised and succinctly summarised their preprocessing methodology, model selection and results. However, I felt that they could add a brief overview of the important features they used could be added to their poster, like they did in their presentation. Poster could also use more graphics (eg exploratory data analysis, or the graph of feature importances, etc). A little more intuitive explanation of the more technical terms behind light GBM on the poster/report would be appreciated.

Presentation-wise, all speakers were clear, fluent and had good pacing. Slides are visually pleasing and easy to follow along.

The group clearly explained motivation and advantages of using light gradient boosting method (LGBM) usage over adaboost and xgboost (bundling mutually exclusive features together.) Good discussion of important features in both the merged and non-merged datasets. However, more varieties of models could be tried.

Quality of writing: 4

Presentation: 4

Creativity: 3.5

Confidence on assessment: 3