Group 3 used Home Credit default risk dataset to predict ability of people to repay loans given historical data. They tried Naive Bayes, Adaboost, linear SGD classifier and Light GBM models.

Report is well written but lacks depth and sufficient visuals (also, text in the correlation heatmap is unreadable.) Exploratory data analysis graphs would be appreciated. In the report, some quantitative results should also be included - for example, in Light Gradient boosting section, what is meant by "learning rate is lower than the default value", "significantly higher amount of estimators than default" "model is computing faster than adaboost"? Some typos and grammatical errors (ex. "worth noted" should be "worth noting") also present.

During presentation, data overview, instead of going feature-by-feature, only described some noteworthy traits about the data and how that would affect their work later on (eg imbalanced class, missing data, curse of dimensionality), which was very appreciated. Clear and concise explanations of different model selections and their respective advantages and disadvantages were given.

Used a variety of model types including Adaboost, SGD classifier, light GBM. Also paid attention to max depth of the model, model can easily overfit. Clearly illustrated the importance of feature selection (removing highly collinear features) on models and how they improved overall result.

Quality of writing: 2.5

Presentation: 4

Creativity: 3.5

Confidence on assessment: 3

=====