Reviewer: Yuan XU

Target: Team 13

Summary of the report.

Team13 predicted the repayment abilities of clients by using machine learning algorithms through information including historical loan records and the characteristics of loan users. To achieve the goal, they use two main datasets. After converting and cleaning data by label encoding and other measures and feature engineering methods. In their work Light GBM model was chosen as the primary model, and for the further test. The feature importance was got in the end.

Describe the strengths of the report.

The report was written in clear format and its structure was built in logistic sequence. As for modeling, Team13 derived some insights from the missing data pattern as people are more likely to skip the optional field when their answers are negative. They regard the missing data pattern as a new indicator and create a new binary indicator called "Incomplete" which will be flagged when there are more and equal to 40 blanks in their applications.

Describe the weaknesses of the report.

A few of typos in the report and poster, like "*% of the total loan amount relative to the applicant's income*". Apparently, there are some numbers missing in this statement. As for modeling, it seems like they have used limited number of variables and parameters, which lead to an ordinary modeling result.

Evaluation on Clarity and quality of writing (1-5):

[4]

The report was written in clear format and its structure was built in logistic sequence. However, a few of typos in the report and poster, like "*% of the total loan amount relative to the applicant's income*". Apparently, there are some numbers missing in this statement.

Evaluation on Technical Quality (1-5):

[4]

The great aspect is that they regard the missing data pattern as a new indicator and create a new binary indicator called "Incomplete" which will be flagged when there are more and equal to 40 blanks in their applications. However, they have used limited number of variables and parameters, and didn't try to use other models base on different theory and compare their performance. And they did not clearly assess both the strengths and weaknesses of their approach in the report, especially for the weakness, missing the part of future improvement in their report.

Overall rating:

4- A good report

Confidence on your assessment (1-3)

3- I have carefully read the paper and checked the results.