Peer Review

Reviewer:

MA Rongyue (Nicole) SID: 20826086

Reviewed Group:

Group 14:

CHAN, Koon Lam

LAM, Chung Wai

TANG, Tsz Hong

1. Summary of the report.

The report contains five major parts - Data preprocessing, Models with different techniques, Evaluation by ROC ad AUC. The logic and structure of the report are quite readable and clear. It really sheds some light on my knowledge of Machine Learning.

Moreover, the final leaderboard score, 0.739, is quite excellent.

2. Describe the strengths of the report.

1. Various selection of model: The group have attempted to use multiple classification algorithms, such as Logistic Regression, Decision Tree, Random Forest, Linear Discriminant Analysis, K-nearest neighbours, Gradient Boosting (XGBoost and LightGBM). Then, they compared the final score of different methods and picked the model with the highest score which is XGBoost. From my point of view, this exhibits their primary thinking process.
2. Well-performed model: The LB score and AUC are all decent.
3. Visualization: There are different graphs for data visualization. Based on those graphs, we can gain a rough view of how data look like, which is very crucial for the following parts.

3. Describe the weaknesses of the report.

The Data preprocessing part can be enriched further. Currently, they only dealt with missing value with “add a column “COMMANAREA_NA” to indicate whether the “COMMANAREA” has missing value)”, and conducted one-hot encoding for categorical values. Maybe, in the further analysis, they can investigate from feature engineering side and try to incorporate some financial domain knowledge.

4. **Evaluation on Clarity and quality of writing (1-5): Is the report clearly written? Is there a good use of examples and figures? Is it well organized? Are there problems with style and grammar? Are there issues with typos, formatting, references, etc.? Please make suggestions to improve the clarity of the paper and provide details of typos.**

   Score: 4

   The report is well-formatted and clear-illustrated.

5. **Evaluation on Technical Quality (1-5): Are the results technically sound? Are there obvious flaws in the reasoning? Are claims well-supported by theoretical analysis or experimental results? Are the experiments well thought out and convincing? Will it be possible for other researchersto replicate these results? Is the evaluation appropriate? Did the authors clearly assess both the strengths and weaknesses of their approach? Are relevant papers cited, discussed, and compared to the presented work?**

   Score: 4.5

   Although there are not many of fancy techniques for data preprocessing and feature engineering, the final score is decent thanks to the powerful XG Boost.

6. **Overall rating: (5- My vote as the best-report. 4- A good report. 3- An average one. 2- below average. 1- a poorly written one)**

Score: 4.5
Overall speaking, the report is good and worth reading.

7. **Confidence on your assessment (1-3) (3- I have carefully read the paper and checked the results, 2- I just browse the paper without checking the details, 1- My assessment can be wrong)**

Score: 2.5 (I have tried to find their Kaggle account but failed, it would be better if they attached the Kaggle account name in the report)