

Peer Review – Report of Group 12

Summary of the report.

The report mainly focuses on the Kaggle competition and works on predicting the clients' repayment abilities using various kind of alternative data. And this report used both the data of training set and some extra data from other institutions and applicants' previous information in this corporation. After data processing and data cleaning, four chosen features were dropped from the Main Table and one extra data was used to fit the model. With using the implementation of Light Gradient Boosting Machine, the fitted model got about 0.75 in both public score and private score with or without the extra data.

Describe the strengths of the report.

The report does well in using extra data including applicants' data from other institutions and applicants' previous information in this corporation. And the fitted model really shows good performance in the testing data set.

Describe the weaknesses of the report.

In the Data Cleaning part of the report, there is too little description about how to get the four dropped features. I think more explanation would be better than only put a link on it.

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

4

The report is in a good format and well organized. And it seems that there is no style and grammar mistake. However, as I have talked about before, in the Data Cleaning part of the report, there is too little description about how to get the four dropped features. I think more explanation would be better than only put a link on it.

Evaluation on Technical Quality (1-5):

5

It seems that the model fitted in Light Gradient Boosting Machine technically works well and shows good performance in the back-testing. I think that there are obvious flaws in the reasoning and the claims are well-supported by theoretical analysis and the back-testing results. And we are possible to replicate the results. Although the author only describes the strengths of the LightGBM approach and does not analyze the potential weaknesses of it, it's still a good report on technical quality.

Overall rating:

4

Confidence on your assessment (1-3)

3