Peer Review

1. Summary:
   The report is divided into five parts, from data description to model validation and selection, to further improvement of the model, and finally proposes future work plans. Through verification, the report finally chose LightGBM among Randomforest, Adaboost, XGBoost, and LightGBM, and carried out a series of further improvement, and finally scored 0.78 in Kaggle.

2. The strengths of the report:
   ① The report is layered clearly, and each part is divided into smaller parts. The content is closely linked, and the logic is progressive.
   ② Dealing with "outliers". The processing of missing data, including the deletion of too much missing data and the reuse of important data (use the feature importance dataframe generated by LGBMclassifier)
   ③ Model selection and promotion. For example, the selection of LightGBM and Dimension reduction.

3. The weaknesses of the report:
   As the author puts forward in the last part, I think there is a lack of application of relevant financial knowledge in the process of data processing. This may cause the existence of some irrelevant data to affect the accuracy of model predictions. At the same time, it will also lead to a lack of interpretability in the choice of variables, which will affect the overall persuasiveness.

4. Evaluation on Clarity and quality of writing: 4

   Typos: Page1:form→from

   Page5:amount→number

   Page7: importance→important

   Some articles use may be incorrect. For example, the author misused the two types of articles 'a' and 'the' in some places.

5. Evaluation on Technical Quality: 4

6. Overall rating: 4

7. Confidence on your assessment: 3