Summary of the report

In this project, they went through the variable selection, model selection and model implementation procedures. Among the four models they evaluate(Logistics Regression, KNN, Random Forest and Gradient Boosting), Gradient Boosting achieved the best performance in cross-validation.

Describe the strengths of the report

The description of feature selection and model selection is very detailed, describing the whole process of model training in an organised manner, with satisfactory final results. A grid search and cross-validation of the hyperparameters used by the algorithm was carried out to improve the robustness of the model.

Describe the weaknesses of the report

Feature engineering is crude. For example, for "NAME_EDUCATION_TYPE", the ordinal numbers are artificially introduced based on subjective judgement rather than using one-hot coding, and the observed outliers are not processed.

Evaluation on Clarity and quality of writing

5 - The format, style and grammar of the report are fine, covering everything from feature selection to model evaluation, and the content is clearly written and organised. Interspersed with images and code to assist with illustration, it is very visual.

Evaluation on Technical Quality

4 - The better final results are attributed to the comparative selection of multiple models, and the rigorous parameter tuning during model selection. However, no relevant papers are cited, discussed and does not elaborate on the reasons for using models for training, which could be supplemented with a comparison of the advantages and disadvantages of different models.

Overall rating

4 - A good report.

Confidence on your assessment

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