Group 10

• Summary

The project applies decision tree-based models (random forest, XGBoost, LightGBM)

on this classification task to predict the loan default risk for every sample. The final

public/private score is 0.77701/0.77984.

• Strength

The data analysis part discusses some basic characteristics of features, which is a

very necessary part of the project. Feature correlation analysis, feature integration

and one hot vector transformation work well on data pre-processing and feature

engineering. The project utilizes different machine learning methods to improve the

score, with several metrics (ROC, f1-score, precision, and recall) in every step listed

clearly in the tables. SHAP value is a good measure of feature importance.

• Weakness

Two figures are very helpful for understanding. It would be better if the figures are

titled and quoted more clearly.

• Evaluation on Clarity and Quality of Writing: 5

The poster is absolutely well organized and formatted with a complete flow of data

analysis. Figures, tables, and formulae are used properly for readers to understand

the detail of data and analysis results. Reference is clear.

• Evaluation on Technical Quality: 5

The project reaches a final public/private score of 0.77701/0.77984, which indicates

the great effectiveness of the models designed by group 10. Also, the discussion

section provides the reasonable analysis for results.

• Overall Rating: 5

• Confidence on Your Assessment: 2