Group 6 review

Summary:

The paper discusses approaches taken in solving the Home Credit Default Challenge. The authors introduced different feature engineering and dimension reduction techniques, including using forward selection efficiently on LightGBM models. The authors used 10-fold cross validation to train LightGBM models, the 10 models are used as members of an ensemble to predict on the test set, they received 0.792 AUC in Kaggle private score. To this end, they claimed that tree-based models are best for this challenge.

Strengths:

The paper is well written and easy to read. The authors made extensive efforts to study the many unique features in this dataset, which had helped readers to understand the meaning behind those features. Dimension reduction is a crucial element in this challenge. The authors accurately noted that usage of linear projection is not adequate to this problem and attempted using UMAP. Forward selection using LightGBM feature importance efficiently is also quite clever. The paper achieved good results and experiments are reproducible.

Things to improve:

We encourage the authors to document quantitative data analysis, the interaction between features/ labels are unclear and could not be explained by feature importance. Forward selection is an important aspect of achieving the results the author claimed, however, the effect of this method on the final score is unclear, an ablation analysis is needed. Finally, it was noted that the combination of UMAP and LightGBM did not achieve satisfactory results, the authors did not attempt to explain this.

Clarity: 5

Technical quality: 4

Overall rating: 4.5

Confidence: 3