1 Group 8

* Summary

Group 8 chose the Home Credit Default as their topic. Choosing four tables with lightgmb model and 5-fold validation. Finally get 77% accuracy. The report is well organized and beautifully made.

* Strengths

The accuracy is very high.

They have tried both generalized models and tree models to do the experiments on the data and made a comparision.

The report is exquisitely made with clear logic.

* Weakness

There seems no EDA for the original dataset.

Some questions:

1. The reasons for table selection and feature extraction may be vague. I have seen the code and haven’t seen the explicit reason for choosing tables and features. For example, report said add some additional features in numerical features but without explicit reason. Maybe losing the analysis of the correlation between different features. In the end, I still couldn't figure out how many features were selected as variables.

2. Report said that they have compared with top100,200,300 features and compared different models but it seems there was not such process in the source code (some of the code have not been run). In the source code it seems there was only one method lightgbm came with the result.

Some suggestions:

1. You can prune the model by optimizing the parameters.
2. You can add more visualization and EDA of the data.

* Evaluation

|  |  |
| --- | --- |
| Evaluation Direction | Score |
| Clarity and quality of writing | 4 |
| Technical quality | 3 |
| Overall Rating | 4 |
| Confidence | 2.5 |

Explicit Reason:

Authors haven’t clearly assessed both strengths and weaknesses of their approach. Some process mentioned in the report haven’t been realized in the code. However, there accuracy is very high and features chose by them are very efficiently and the report is clear and well organized.