READ\_ME

Overview:

This competition is held by American Express which is a globally integrated payments company and also the largest payment card issuer in the world. In this competition, we could apply our machine learning skills to predict credit default. Specifically, we will leverage an industrial-scale data set to build a machine-learning model that challenges the current model in production. Training, validation, and testing datasets include time-series behavioural data and anonymized customer profile information. We can explore any technique to create the most powerful model, from creating features to using the data in a more organic way within a model.

Notice:

1.Raw dataset is in the path: input/amex-default-prediction

2.Pre-processed dataset is in the path: input/amex-data-integer-dtypes-parquet-format OR (https://www.kaggle.com/datasets/raddar/amex-data-integer-dtypes-parquet-format)

3.Code/fe\_process.py this process aims to generate new features, which would take less than an hour.

4.Code/lgb.py This is for training the Lgbm model

5.Code/xbg.py This is for training Xgboost(Use GPU)

6.Code/lgb\_2.ipynb This is for training the 2nd Lgbm model

7.Code/infer.ipynb This is for getting the final output.

Libraries needed:

Pandas

Numpy

Lightgbm

Pyarrow

Pickle

Tdqm

This competition needs high Ram(prefer 64GB)