ML Model Assessment Case Study

A credit card fraud model is trained on a credit card transaction dataset using XGBoost to prevent transaction fraud. The model output is a probability and any transaction that has probability>0.5 will be rejected due to fraud risk.

Please find 3 attachments related to the model:

- creditcard_interview.csv: dataset used to train the model
- Interview_Case_Code.ipynb: notebook used to train the model
- interview_model_object: final model object (you can find the codes to load the model object in the notebook Interview_Case_Code.ipynb)

1. Please provide your assessment of the overall model development process as well as the selected model, and discuss potential issues, limitations, opportunities for improvement, if any.

2. If the company loses 100% on the fraud transaction, and earns 2% transaction fee on any non-fraud transaction, describe how you are going to use this information in the model validation.

3. Please build a benchmark model using coding languages such as Python and perform a comparison against the XGBoost model provided. The benchmark model is not expected to exactly replicate the XGBoost model - please try to use a better modeling approach where you see fit and make your own assumptions.

Note: Please discuss your approach and thought process as detailed as appropriate.