This is financial fraud data about a bank and payments system. You will have to use this data to predict whether the transaction is Fraud or not ( isFraud column)

There are 6 million records in this dataset.

You can use first 4 million records for building the model, next 1 million records for evaluating the model and rest of the records as live records on which model would be evaluated in production.

Please create separate .py files for model training and prediction.

Make sure that your dataset during model building and prediction phase goes through same data processing and feature creation steps.

Your final deliverable can either be a python script which can be scheduled as a batch script or flask api endpoint which can be used to identify fraud transactions in real time.

Please submit the final report which give clear idea on below points.

1. ML model selected.
2. Features selected.
3. Feature importance
4. Evaluation of model
5. Original code

Use the “Fraud\_detection” table from the Database.db file for the project on predicting fraud

transaction.

Please feel free to reach out to your mentor in case of any doubts.