The complete prediction is divided into 3 python scripts:

1: **Data\_Preprocessing.py**: This script outputs the needed dataset to be able to feed into the model by using the dataset given after generating it from the servers (here it’s like the time\_between\_two\_tolls.csv ) and saves the required output as ‘*Processed\_df.csv*’

2: **Model\_py**: This script uses this ‘Processed\_df.csv’ as input to train the model with the optimum hyperparameters saves the trained model as ‘*finalized\_model.sav*’ using the pickle package python. The other files that are pickled are the list of vehicle number, toll\_both number assignment dictionaries, as we need to convert the test dataset toll\_both, vehicle\_no fields in the exact same numbering system as we did with the train\_df. This scripts takes some time as it trains the model.

3: **Prediction\_py**: This script is used to supply the test data and use it to predict the time of arrival for the remaining tolls using the saved model ‘*finalized\_model.sav*’. The predicted results of the forthcoming toll\_boths are saved in the csv file of ‘*predicted\_eta.csv*’. This script takes practically negligible amount of time as it just predicts using the already trained model.