**svm.py**, **rf.py** and **xgb.py** were used to conduct the hyper-parameters optimization based on the data folds from Attentive FP and the subsequent 50 times independent runs for SVM, RF and XGBoost respectively.

**dnn\_torch.py** and **dnn\_torch\_utils.py** were used to conduct the hyper-parameters optimization based on the data folds from Attentive FP and the subsequent 50 times independent runs for DNN.

**gnn.py** and **gnn\_utils.py** were used to conduct the hyper-parameters optimization based on the data folds from Attentive FP and the subsequent 50 times independent runs for four graph-based model include GCN, GAT, MPNN and Attentive FP.

**data\_wash.py** was used to conduct data washing as described in the “Washing of the Benchmark Datasets” section.

All the scripts were ran in a Linux server installed the following main packages:

Python (Version: 3.6.5 x64)

PyTorch package (Version: 1.3.1+cu92)

scikit-learn package (Version: 0.20.1)

XGBoost (Version: 0.80)

DGL package (Version: 0.4.1)

RDKit package (Version 2019.09.1)

hyperopt package (Version: 0.2)

Linux MOE (Version 2015.1001)