**The explanation of the data used to train**

**Zihao Liu**

1. I have already downloaded all the available original data of Electricity, ETTm1, Mujoco, PTB-XL248, PTB-XL1000 from their original Github repo. This data is the raw data without the feature\_sample\_processing provided by the paper authors.
2. The paper authors also provide a quick way to download the processed available data through “get\_data.ipynb”. The data in this file is all from the “/content/drive/MyDrive/JPMorgan/自写代码/Necessary\_codes\_from\_original\_authors/get\_data.ipynb”
3. If we run the codes in “get\_data.py” then we will obtain the directly available both train and test datasets. So we just run that code and get the data which are after processed by the authors. Then we can use “train.py” to train them.
4. If we need to see the very original data, we can find them in “/content/drive/MyDrive/JPMorgan/SSSD-main/docs/instructions” In these specific data files, the link to the original data is provided. There is also the “feature\_sample\_process.ipynb” to process these original data. It would be relatively hard for some kinds of data to be processed from the original data. It is why we just use “get\_data.py” to directly obtain the processed data provided by the authors.