**Journal Report 12**

**11/10/23**

I had a problem reading in the pickled dataset in kaggle. It seems that the data was corrupted or had some kind of formatting issue when I uploaded it into Kaggle. I fixed the problem by directly accessing the CCR dataset and resampling each time series. This took an hour because I had some dependency issues regarding the Darts library and Darts time series object. I fixed these problems by downgrading, but Kaggle seemed to be very slow on this day. By the end of the class, I managed to recreate my environment in Kaggle.

**11/13/23**

I was researching different implementations and libraries for creating synthetic data over the weekend because the current implementation relies on an outdated library. I had to downgrade and create a new anaconda environment to get the library to work on my personal computer. However, I cannot do this on Kaggle so I thought the best course of action would be to find a new implementation.

I found a library called ydata\_synthetic, which has an interfaceable GUI for creating synthetic data from a dataset. It's a python library that provides a collection of generative models for easily creating high-quality synthetic data for applications like augmenting small datasets, protecting sensitive data, and exploring new data scenarios. It supports various models including GANs, VAEs, and Gaussian Mixtures, and offers a user-friendly API, pre-trained models, and custom model training capabilities.

It seems to be user-friendly with good documentation, but I have yet to see how it performs. I hope to implement it by next class.

<https://github.com/ydataai/ydata-synthetic>

**11/15/23**

Today, I spent my time implementing the TimeGAN model from the ydata\_synthetic library. I had another problem with importing the libraries, but it turned out to be a quick fix. I used the same method I used last week by downgrading my library to a specific version stated in Stack Overflow.

I’m still figuring out how the library works. I implemented the code to train the model and preprocess my data so that it can be fed into this new GAN implementation. I don’t quite understand the different parameters set for the model as I am using the default ones given by a tutorial:

<https://towardsdatascience.com/synthetic-time-series-data-a-gan-approach-869a984f2239>

gan\_args = ModelParameters(batch\_size=128,

lr=5e-4,

noise\_dim=32,

layers\_dim=128,

latent\_dim=24,

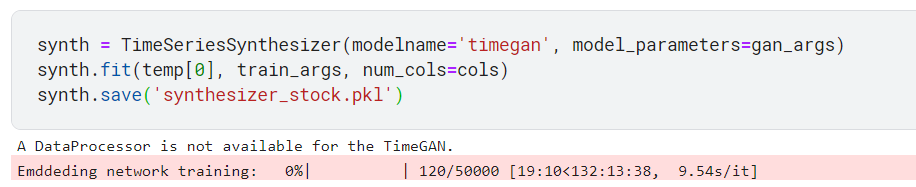
gamma=1)

train\_args = TrainParameters(epochs=50000,

sequence\_length=24,

number\_sequences=12)

I get this error when training:



I think that it has to do with the parameters so I will research more into that next class.