WADI code file name (Note: variables adjusted during algorithm assessment and code may not run without changing variable names. Best to run cell by cell to adjust variables as needed. Required packages and libraries listed in cells)

- KMEANS_WADI, (KMeans.ipynb, KMeans2.ipynb, KMeans3.ipynb)– novelty detection, subsystem grouping on WADI data set
    - Ensure that "WADI_14days_new.csv" and "WADI_attackdataLABLE.csv " are available.
- SWAT_2015_CUSUM (SWAT_2015_CUSUM.ipynb) – CUSUM anomaly detection on SWAT 2015
    - Ensure that "SWaT_Dataset_Attack_v0.csv" is available. Initial code for KMeans not used for project.
- LSTM_KNN (EDA_LSTM.ipynb, LSTM.ipynb, LSTM2.ipynb, PyOD.ipynb, PyOD2.ipynb, PyOD3.ipynb) - LSTM and KNN anomaly detection WADI (EDA included)
    - Ensure that "WADI_14days_new.csv" and "WADI_attackdataLABLE.csv " are available.

main_Clustering.py

Usage: main_Clustering.py

The main_Clustering.py will run on SWaT 2015 and SWaT 2019 datasets with the below inputs and variable changes. The script is designed to run single or numerous test cases and produce visualization and metrics to assess the performance of the PCA and K-Means clustering outlier detection implementation.

Input:

- Modify the "test_cases_str" variable to a .csv with desired test cases to execute. The Columns and rows of the .csv as formatted as follows:

Case	clusters	PCA_Components	Threshold
1	5	5	2.5
2	6	5	2.5
3	7	5	2.5

- Modify "sub_folder" variable to directory where data file exists.
    - Example: "/SWAt2019/"
- Modify "file_name" variable to specific data file in directory above
    - Example: "SWaT_Dataset_2019.csv"
- Modify "trial" variable to specify if "test case" functions will run vs. All functions will run
    - If trial = True, an variable labeled "options" is set to equal an array as follows: [False, False, False, False, False, True, False]
        - This will execute only the confusion matrix scripts to grade the accuracy of all test cases
    - If Trial = False, the "options" is set to equal [True, True, True, True, True, True, True, True], indicating all functions and below plots will be produced.

Output:

- For each True/False value in the options array, the below outputs will be produced:

Index	Option array value	Output
0	True	Variance vs. PCA plot, PCA 1 vs. PCA 2 plot
1	True	Cluster vs. Inertia Plot (Elbow Plot)
2	True	PCA 1 vs. PCA 2 plot, PCA 2 vs. PCA 3 plot seaborn plots
3	True	3D Scatter Plot of clusters out of K-Means Model
4	True	3D Scatter Plot of original Normal vs. Attack data points/labels
5	True	Execute grade_Outlier function that will create confusion matrix with TP,TN,FP,FN and accuracy metrics
6	True	Produce Results.csv with confusion matrix and accuracy metrics

“SWaTJUL2019_EDA_CUSUM_K-means_match.ipynb”

- Ensure that you have both "SWaT_dataset_Jul 19 v3 (edited).csv" and “SWaT_clusters_2019.csv” loaded in the same working directory as this file
- Press “run all” to view dataframe clean up, normalization/standardization, early data analysis, CUSUM for SWaT2019, and CUSUM + K-means time matching for SWaT2019

“SWaTDEC2015_CUSUM_K-Mean_match.ipynb"

- Ensure that you have "SWaT_Dataset_Attack_v0.csv", "SWaT_Dataset_Normal_v0.csv", and "Case_0_SWaT2015_P1_clusters.csv" loaded in the same working directory as this file
- Press “run all” to view CUSUM for SWaT2015, and CUSUM + K-means time matching for SWaT2015