**\*\*Project 4 (Big Data Applications)\*\***

**Project Overview:**

* This project aims to process real-world sensor measurements for machine learning robot force prediction model.
* We have 12 sets of asynchronous multivariate time series dataset from 6 different sensors which should be processed into a single data matrix.
* The dataset also includes sensor bias, variance shifts, noises, outliers etc those should be cleansed and corrected.

**Presentation and Q&A:**

* **Submission and Presentation Deadline:** To be decided in class.
* **Presentation Duration:** 15 – 20 minutes.
* **Q&A Session:** 10 – 15 minutes following the presentation. Each team member must be prepared to answer questions individually.

**Project Goals:**

* To synchronize all asynchronous multivariate time series into a single matrix.
* Eliminate experiment-wise bias, variance, noises, and outliers.
* Perform trend-residual decomposition using Fourier transformation.

**Project Tasks:**

1. **Time vectors synchronization:** Interpolate in time and synchronize time vectors to create data matrix for machine learning computations, with minimum arbitrary value estimations for missing time steps.
2. **Sensor bias correction:** Remove experiment-wise initial bias within each sensor to stabilize training convergence.
3. **Noise filtering:** Eliminate exceedingly high-frequency components from measured time series, which may hinder the model’s prediction by Fourier filtering.
4. **Outlier removal:** Cut-out invalid measurements, such as negative pressures or anomalies with appropriate method.
5. **Sensor coordinate unification:** Unify different sensor coordinates into global coordinate.
6. **Trend-Residual decomposition:** Extract trend using Fourier transformation and residuals (remainders) by summation decomposition.
7. **Residual variance correction:** Correct serious shifts in residual series variances along experiments.
8. **Visualization:** For all above processes.
9. Conduct a trend analysis on your data. Then, eliminate 10% of the data in a random pattern and fill in the blanks with any statistics of your choice (mean values of the rest, ...). Repeat the trend analysis and compare the results with the results of the original data. Remind that all tasks should be visualized.

**Project Allocation:**

* This project will be assigned to the first group that applies.
* To apply, email me as soon as possible.

**Good Luck!**