PROJECT 1

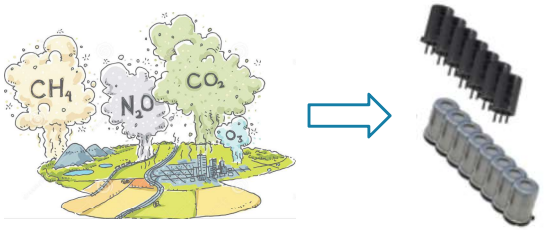
CO CONCENTRATION LEVEL PREDICTION

8조 진형남, 박민수, 신명진, 장대근

목차

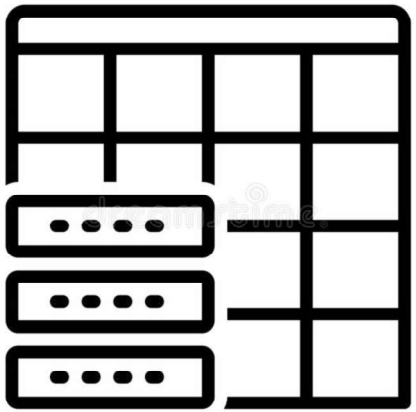
- Problem Statement
- Exploration
  - ✓ Domain Exploration (Literature Review)
  - ✓ Data Exploration
- Methods & Results
  - ✓ Overall Approach
  - ✓ Summary: Models & Results
- Conclusion

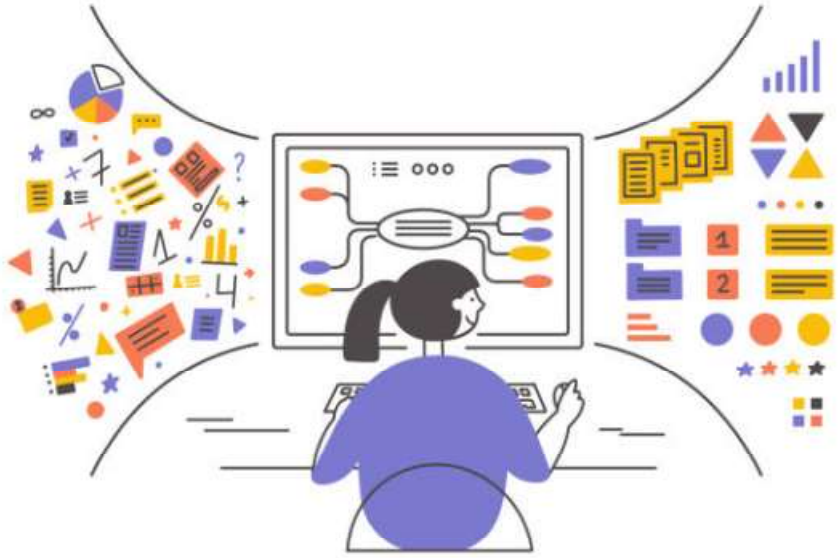
Problem Statement



Predict CO concentration levels from 16 gas sensor data

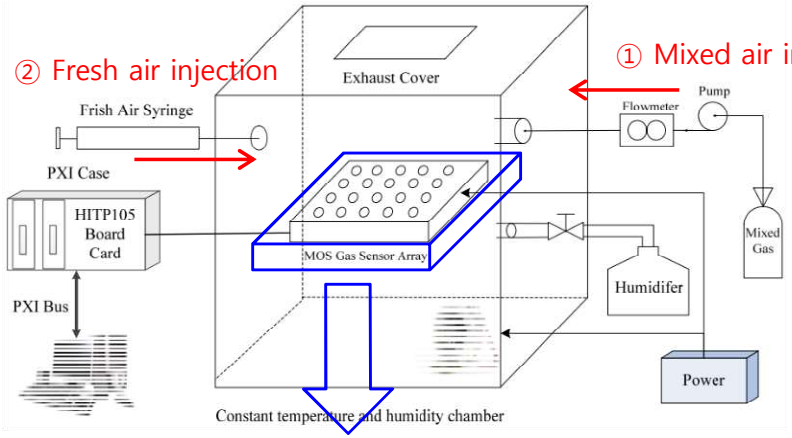
Given data

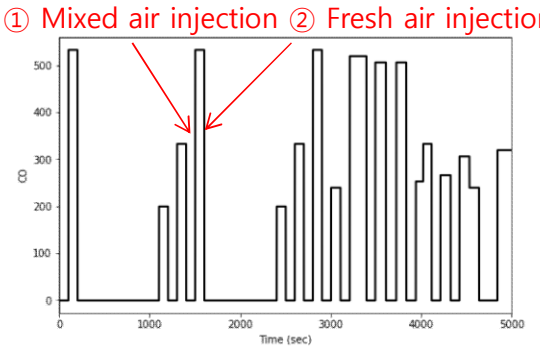




Domain Exploration (1/2) by Literature Review

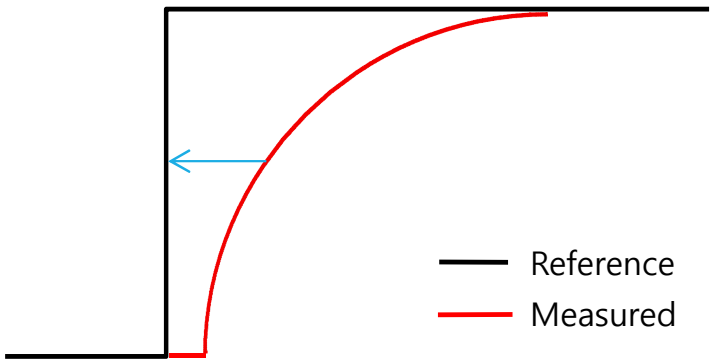
- (General) In-Lab Mixed Gas Test Environment





- Delayed sensor response due to:
  1) Remaining gas in the chamber (before fully exhausted)
  2) Sensor array structure
  3) Sensor internal characteristics (∵ analog lowpass filter)

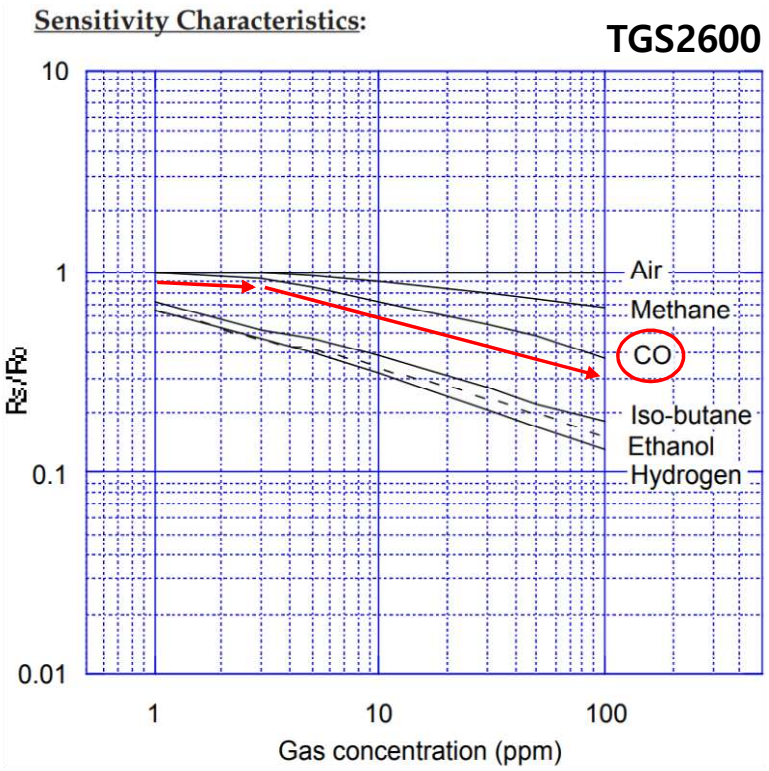
☞ **Need to align measured signals with reference CO**



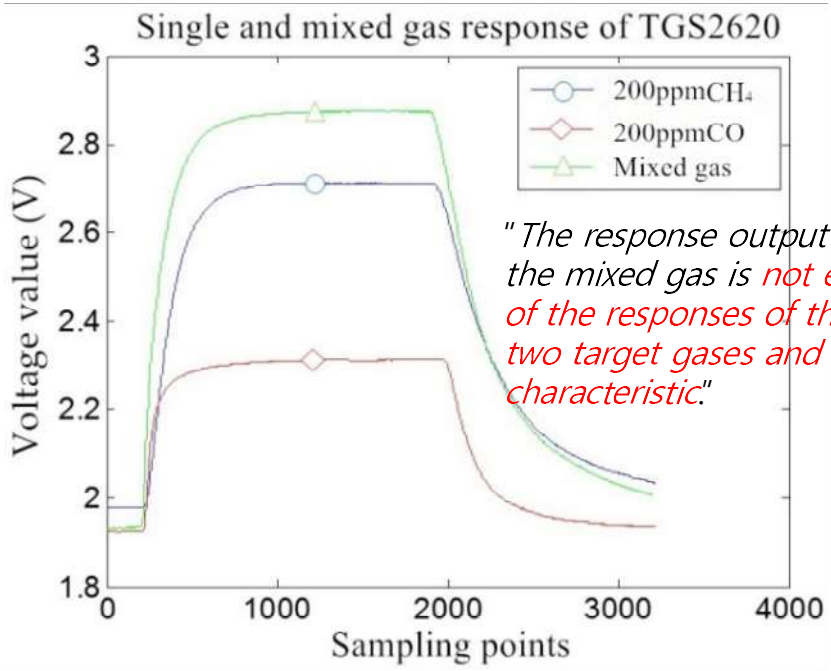
*Ref.: Yonghui Xu, et al., "Research on a Mixed Gas Recognition and Concentration Detection Algorithm Based on a Metal Oxide Semiconductor Olfactory System Sensor Array," *Sensors*, 2018

Domain Exploration (2/2) by Literature Review

- Nonlinearity in Cross-sensitivity characteristics & Mixed response



☞ **Model should represent these nonlinearities!**



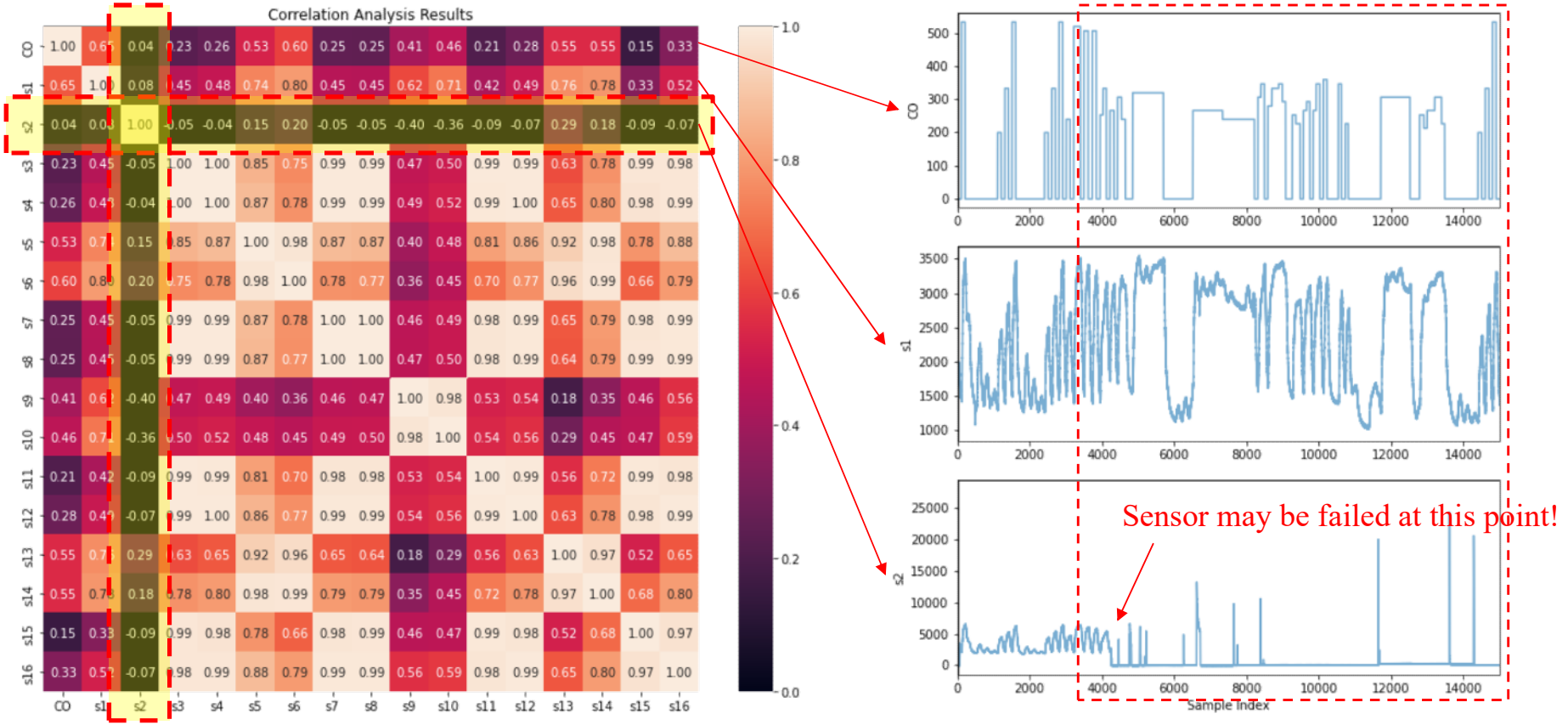
"The response output of the sensor to the mixed gas is not equal to the sum of the responses of the sensor to the two target gases and has a nonlinear characteristic."

*Ref.: Datasheets (Sensor specifications) of FIGARO TGS Series Sensors

*Ref.: Yonghui Xu, et al., "Research on a Mixed Gas Recognition and Concentration Detection Algorithm Based on a Metal Oxide Semiconductor Olfactory System Sensor Array," *Sensors*, 2018

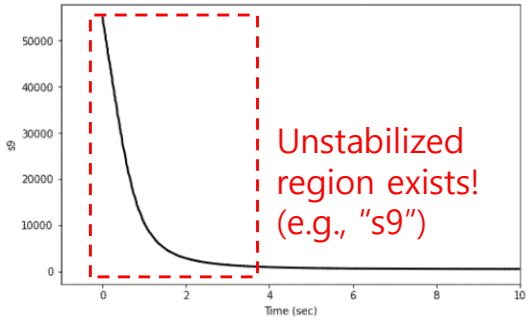
Data Exploration (1/3)

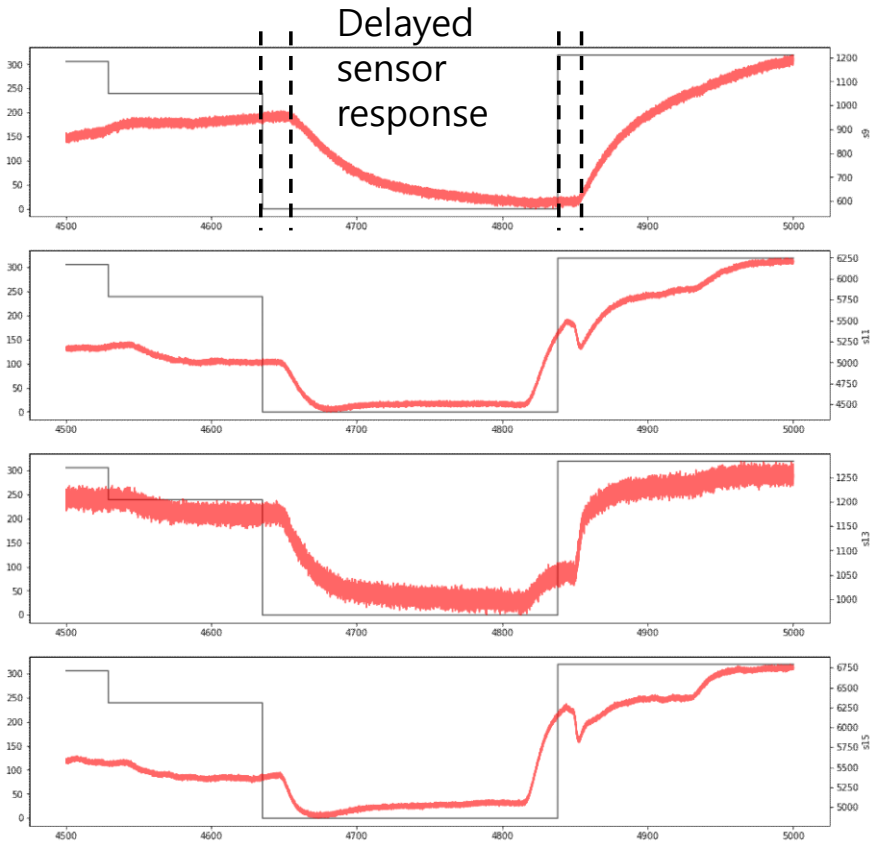
- Cross-Correlation Analysis: ① Sensor Failure @ "s2" → Exclusion?
  ② Similar Sensors → Selection or Merge?

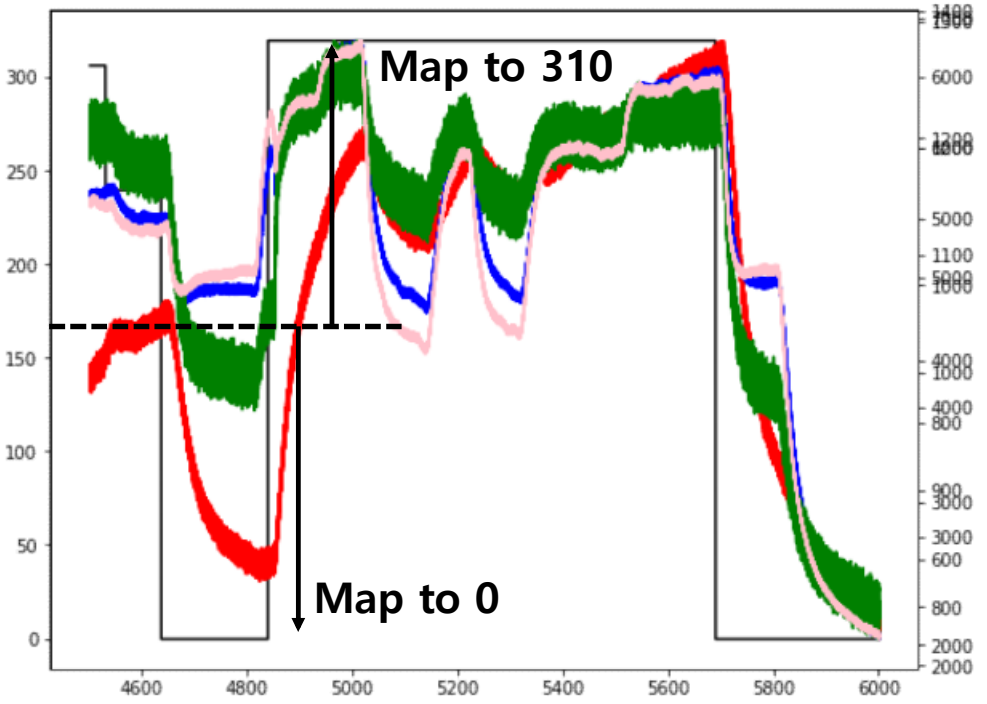


Data Exploration (2/3)

- Need of Signal Shifts → **2000 samples shifted!**



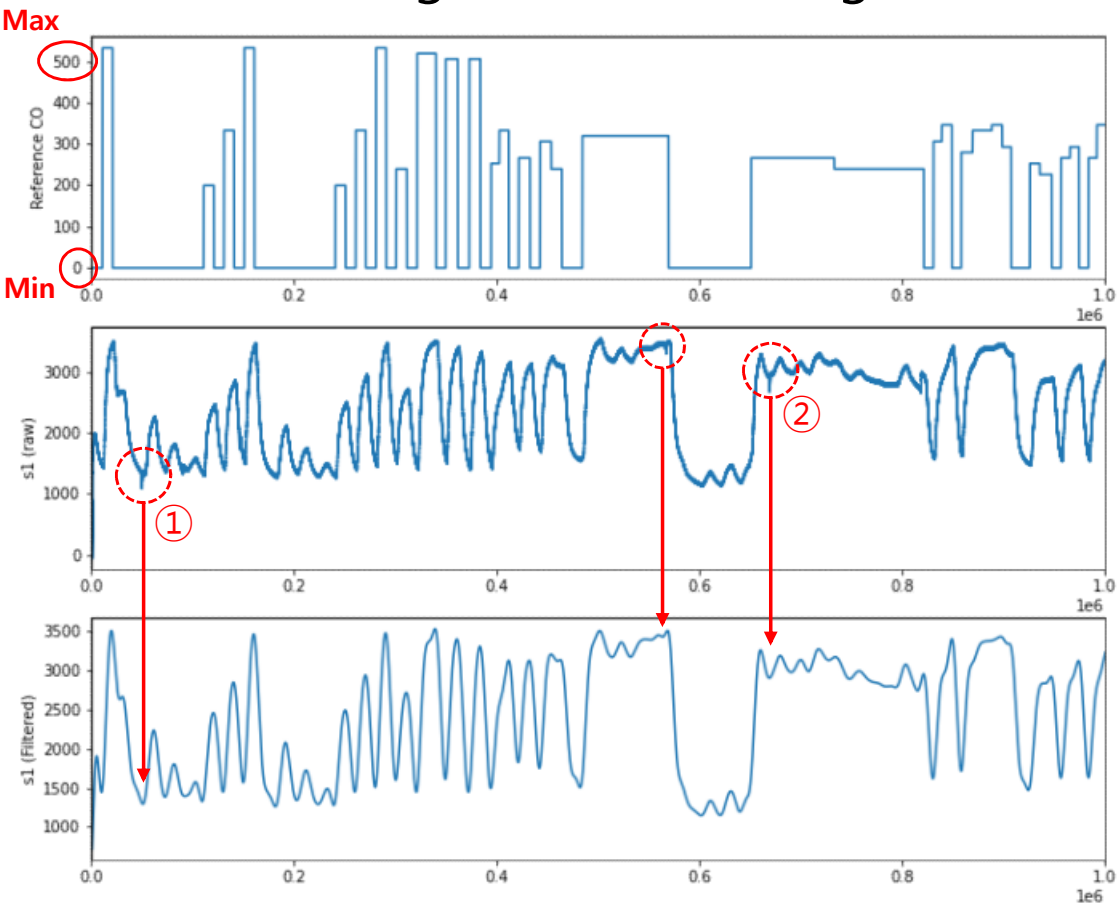


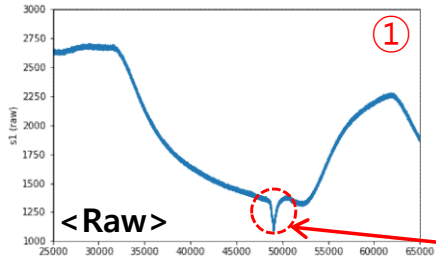


*Measured signal = Lowpass-filtered form of target signal
**Poor ability to track abrupt change in CO concentration level

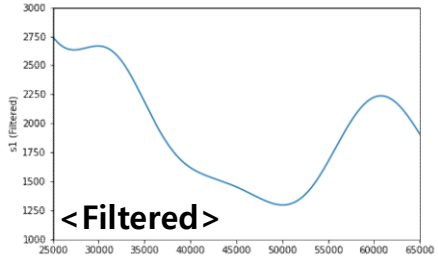
Data Exploration (3/3)

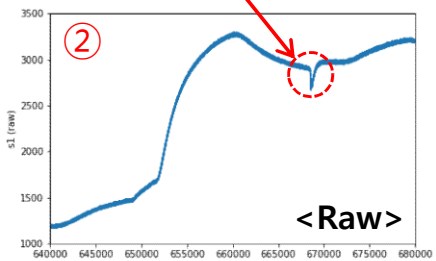
- Need of Signal Conditioning → **Zero-Phase Lowpass Filter Applied!**

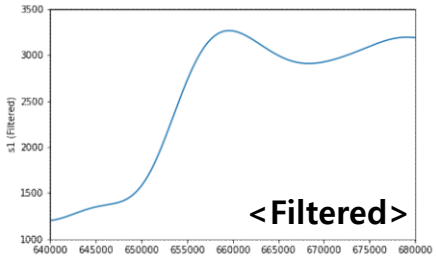




Impulsive noise exists!

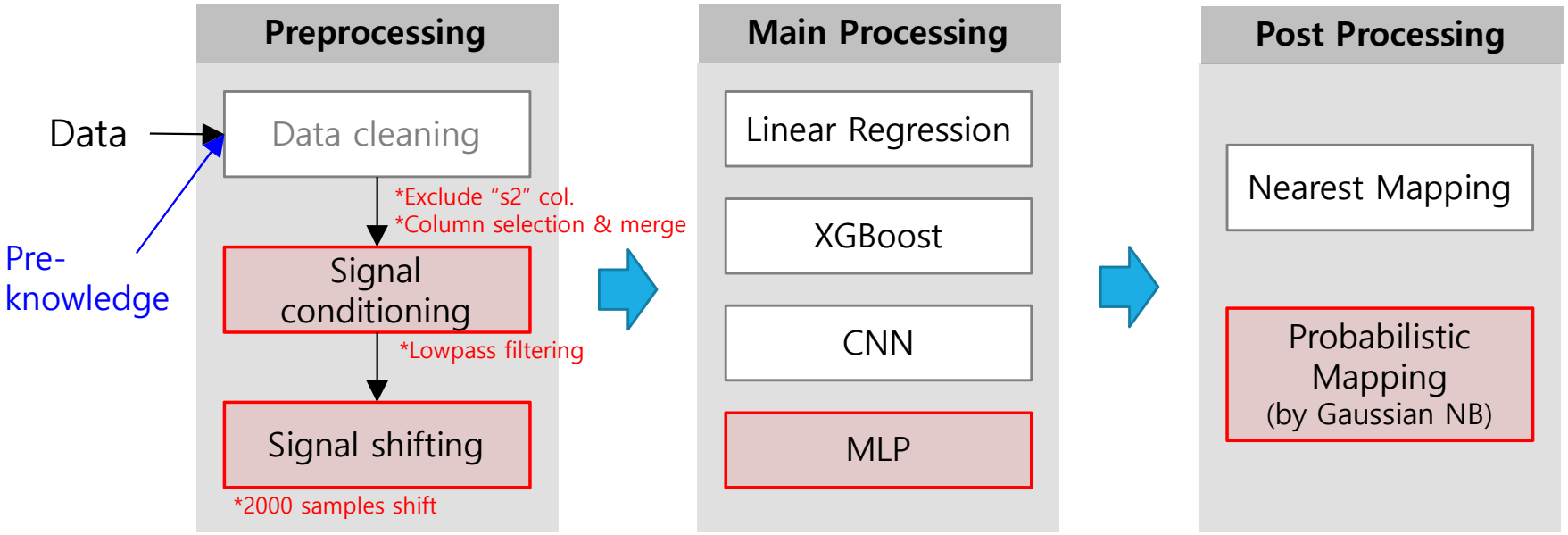


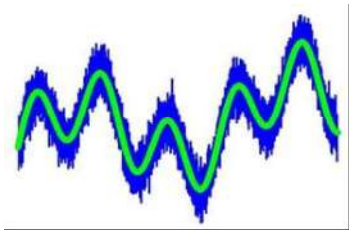


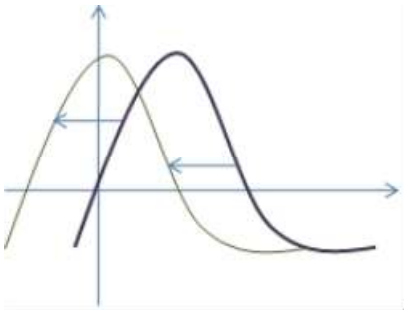


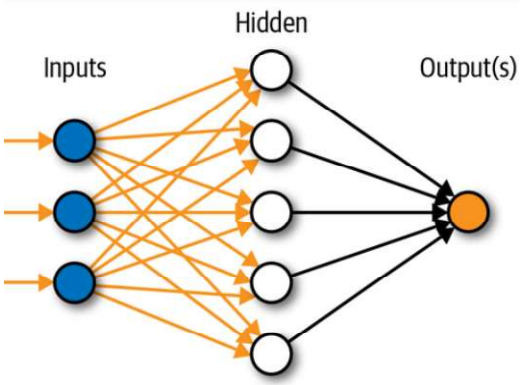
Overall Approach

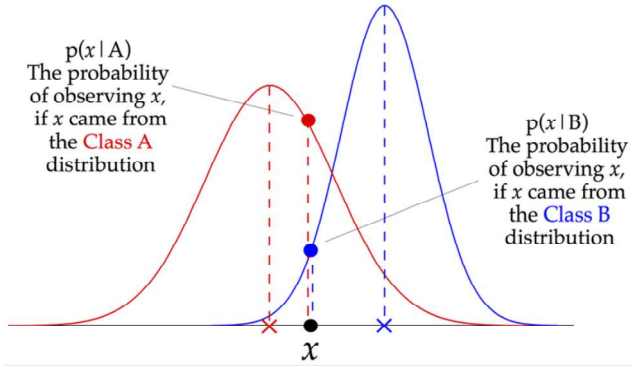
Best strategy



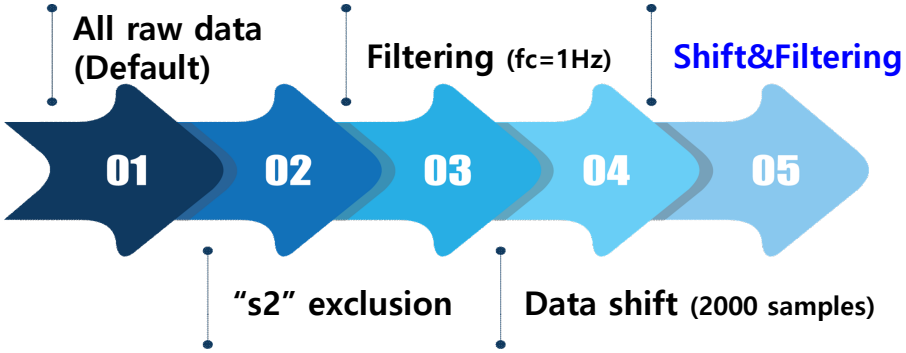


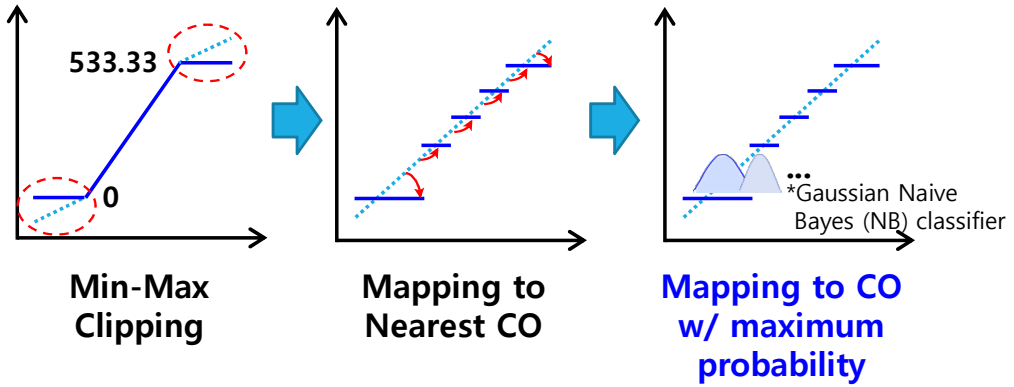


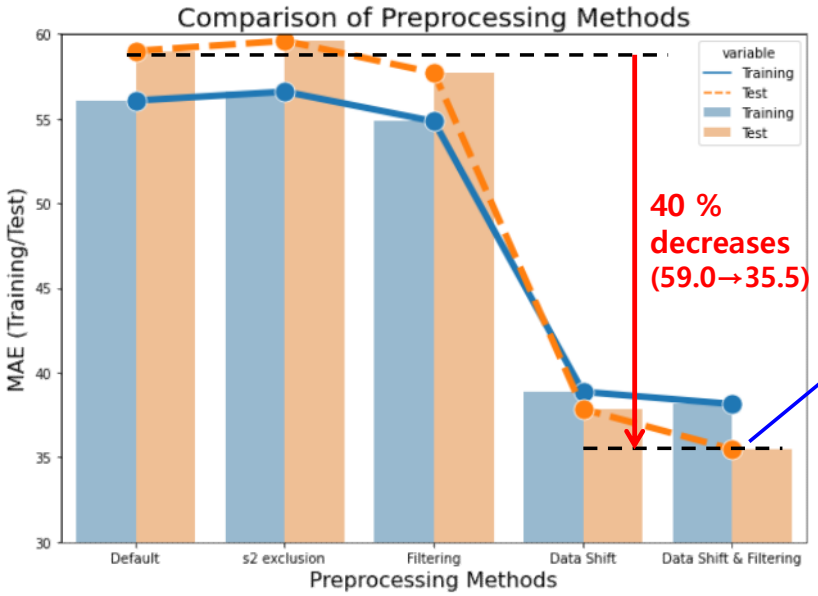


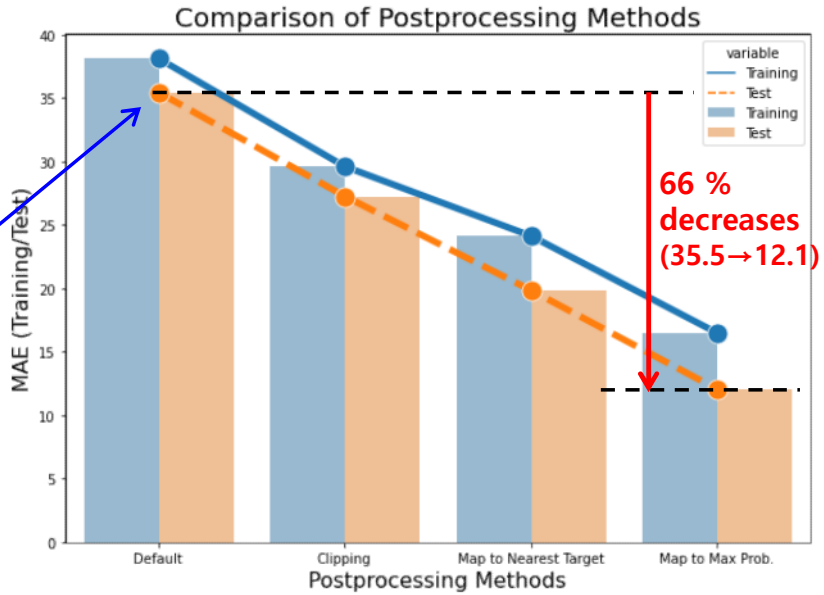


Pre- & Post-processing (79.5 % MAE reduced w/ LR)

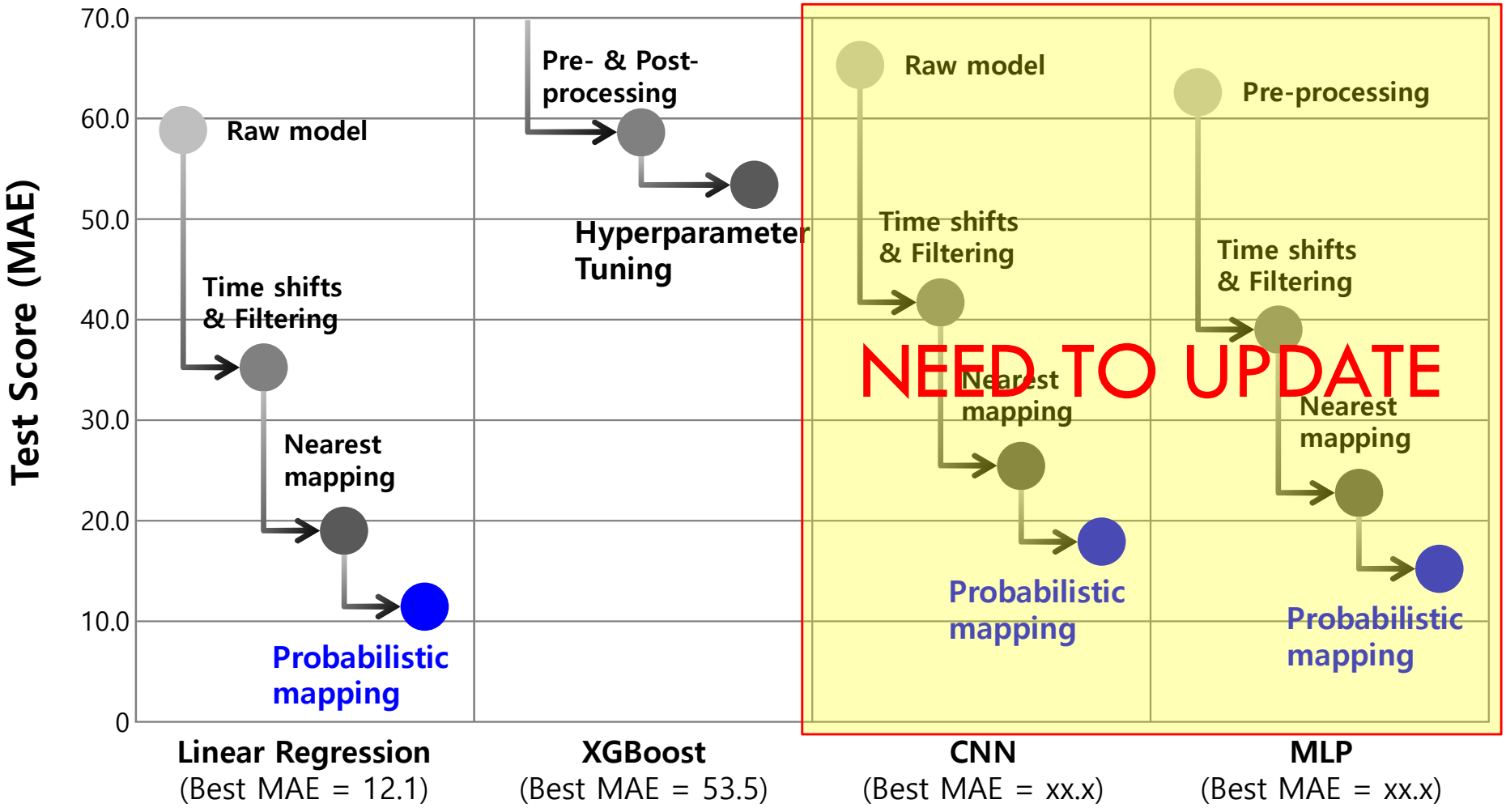








Summary: (Combined) Models & Results



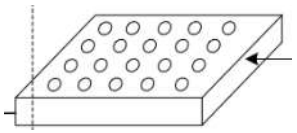
Conclusion

"No Free Lunch"

- **Linear regression**
  - ✓ Simple but powerful (∵ highly correlated data)
  - ✓ Limited performance (∵ only linear relationship)
- **XGBoost**
  - ✓ Good for classification (than regression tasks)
  - ✓ Poor at extrapolation (depends on training set)
- **CNN**
  - ✓ Good for data with spatial relationship
- **MLP - Achieved the best performance**
  - ✓ Good for time-series prediction (w/ tabular data)
  - ✓ But too many parameters (∵ full connection)

Further Works

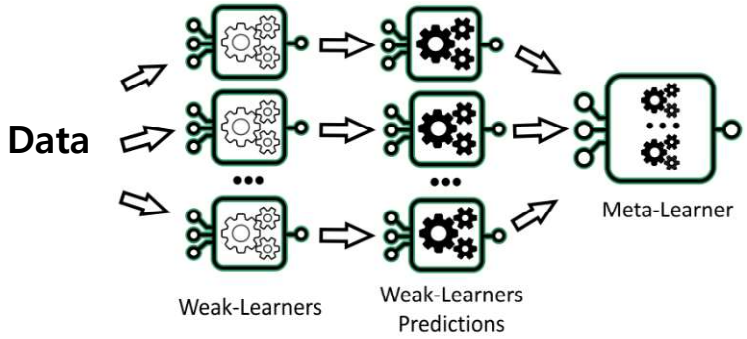
- **Different time shifts per sensors (columns)**



	s1	s11	s13	s15
Optimal Time Shift*	4700	2700	2300	2200

*determined by the maximum cross-correlation

- **Use CNN w/ spatial representation** (e.g., spectrum)
- **Use RNN models - e.g., LSTM, Transformer**
- **Ensemble model**



*Ref.: Towards Data Science, "Why XGBoost Can't Solve All Your Problems"
*Ref.: MachineLearningMastery, "When to use MLP, CNN, and RNN neural networks"