

23.10.26.

다변량 스팀 사용 이상 감지 및 영향 변수의 원인 분석

서울과학기술대학교 산업공학과/데이터사이언스학과

이성호 sean0310@seoultech.ac.kr

배소희 shbae2819@g.seoultech.ac.kr

심재웅 jaewoong@seoultech.ac.kr

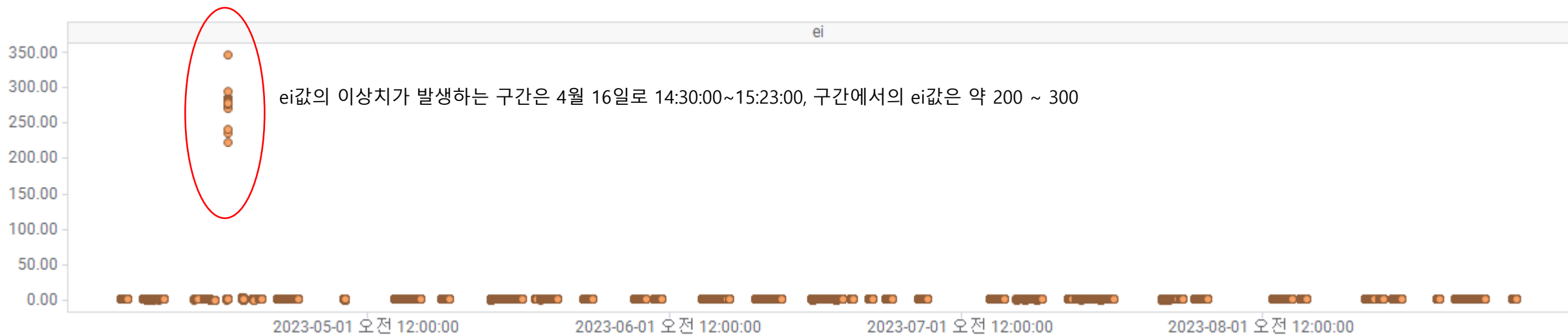
Contents

- 데이터
 - 데이터 이해
 - 실험 데이터 전처리
- 모델링
 - 머신러닝 모델 (**Linear regression, Random forest**)
 - 딥러닝 모델 (**1D CNN, LSTM**)
- 향후 계획

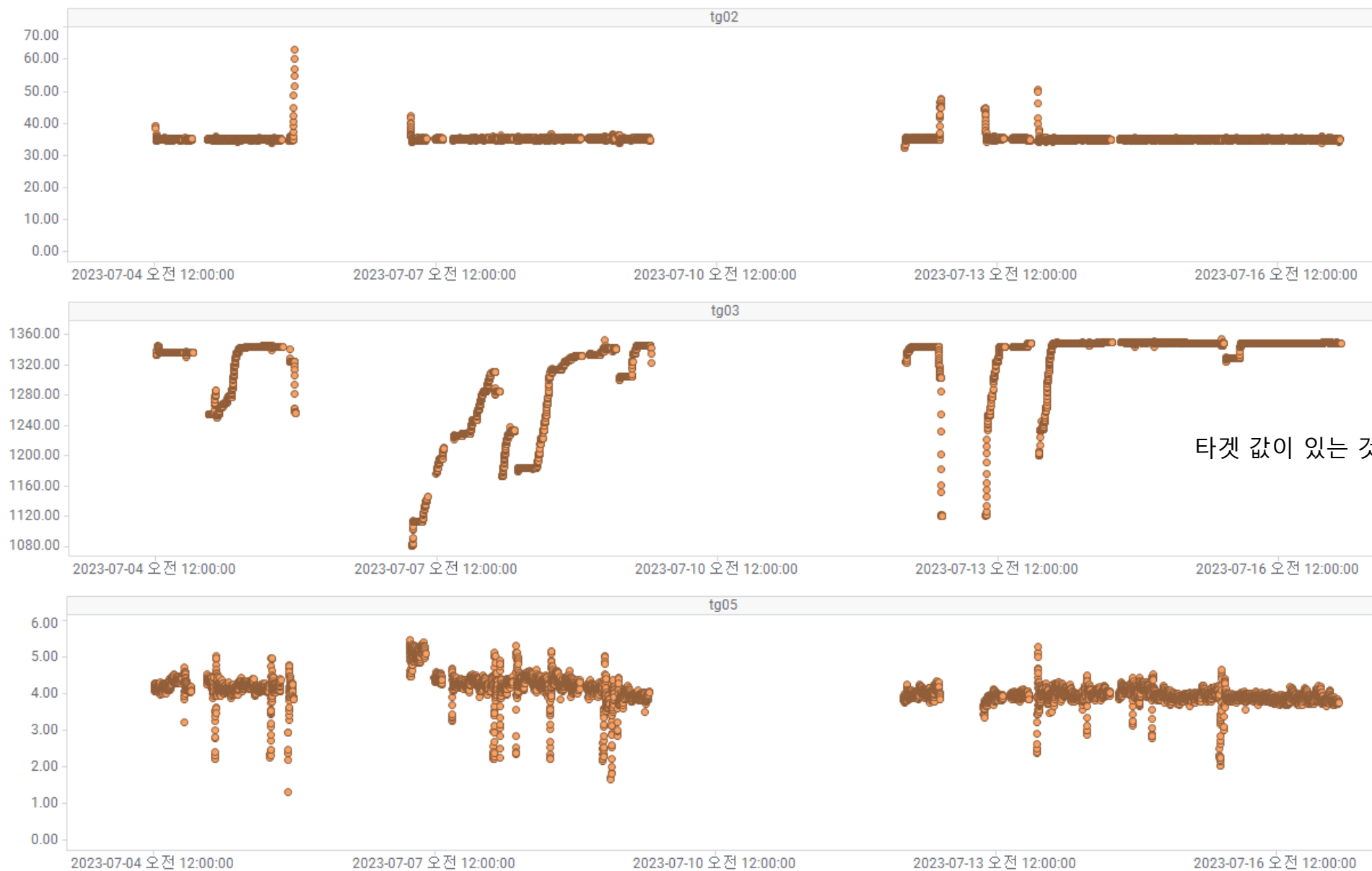
데이터 이해

■ 이용 데이터

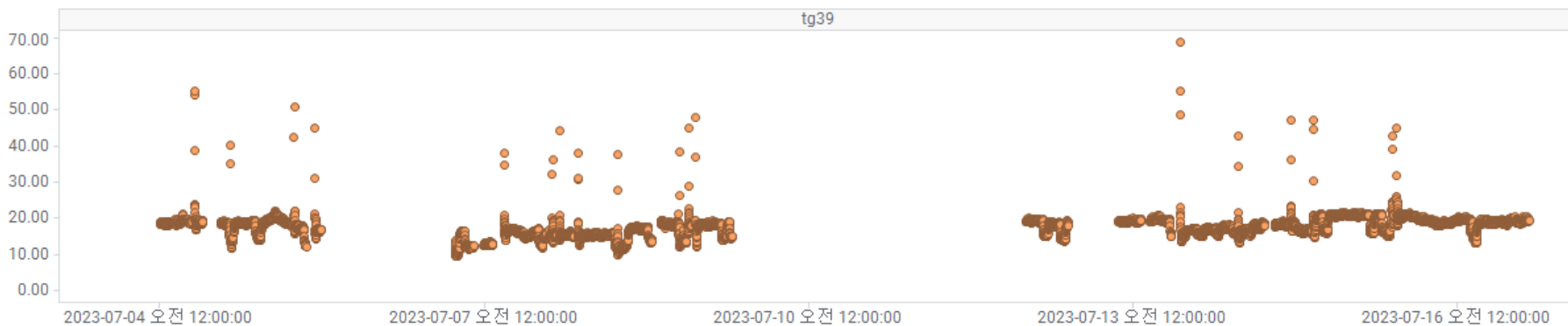
- `df_ext(2023-04-01~2023-08-31,51250385)_2023-10-17 10-58-30 -seoultec` (제품 1종에 대한 한달간 센서 데이터)
- 제품 코드 : 51230385
- 기간 : 2023-04-05 14:59:00 ~ 2023-08-27 03:00:00 (분)



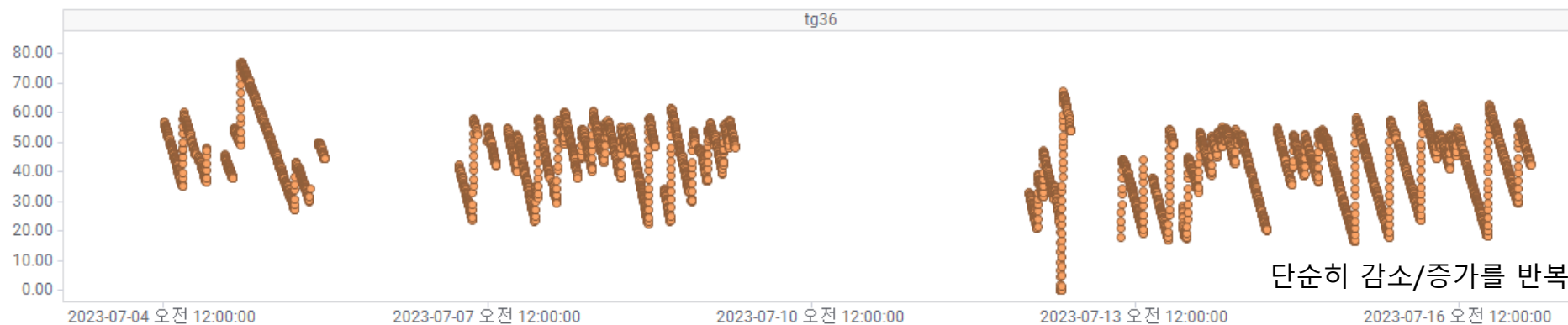
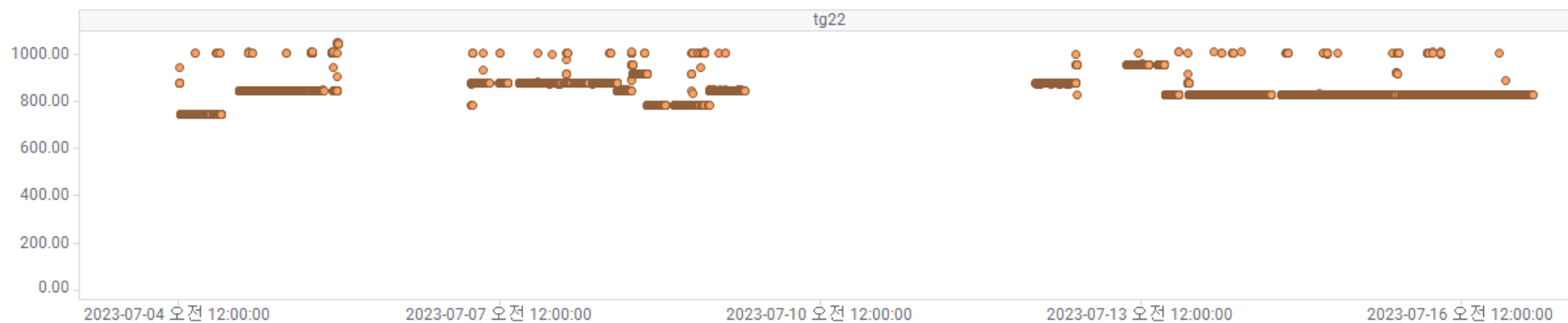
데이터 이해



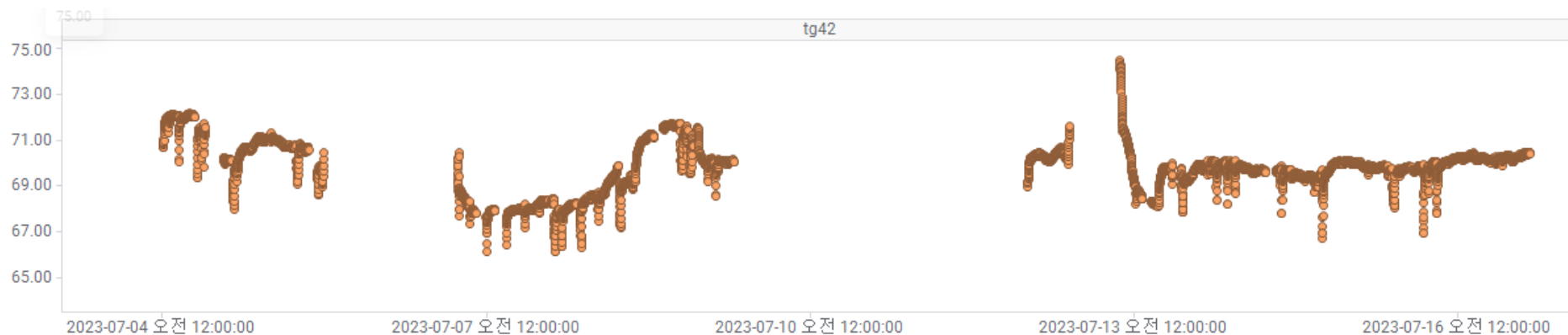
데이터 이해



데이터 이해



단순히 감소/증가를 반복하는 값



실험 데이터 전처리

■ 이용 데이터 전처리

- output : ei와 중복이기 때문에 제거
- sstable : 동일값(1)이기 때문에 제거
- jr : 생산품의 번호 제거
- shift, wclass : 카테고리 변수로 제거
- stop : 정상가동으로 동일값(0)이기 때문에 제거

■ Data split

- Stop 및 다른 이유로 인해 끊긴 시간대 지점들을 고려하여 다음과 같은 비율로 분할
- Train : Test = 0.8 : 0.2 (딥러닝 모델의 경우 train set의 일부를 early stopping에 사용)

■ 예측 모델 구성

- Regression 모델
 - **input(X)** : 과거 7분 동안의 38개 sensor값
 - 끊긴 시간대로 인해 잘려서 기록된 데이터의 최소 크기가 7(2023-04-13 02:53:00 ~ 02:59:00)이기 때문
 - **output(y)** : ei값

date	날짜
tag	Sensor (38개)
output	제품 생산량 계산값(현재)
ei	원단위 계산값(=스팀 사용량/output)
sstable	원단위 상태 분석값 0:잘음 1:적당 2:나쁜
jr	단위 공정값 / 제품 생산 주기 (생산품 번호)
shift	작업팀 구분값
wclass	작업팀 구분값
stop	공정 분석값 0:가동 1:중지-이벤트-발생 2:중지-복구

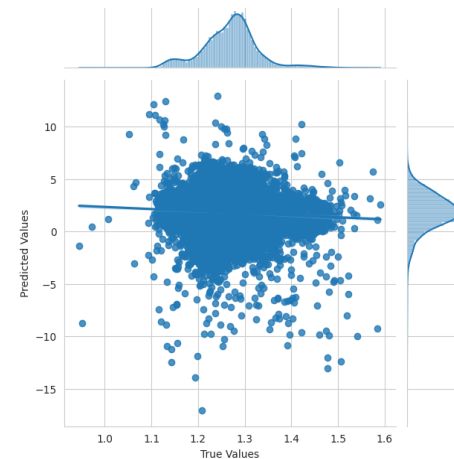
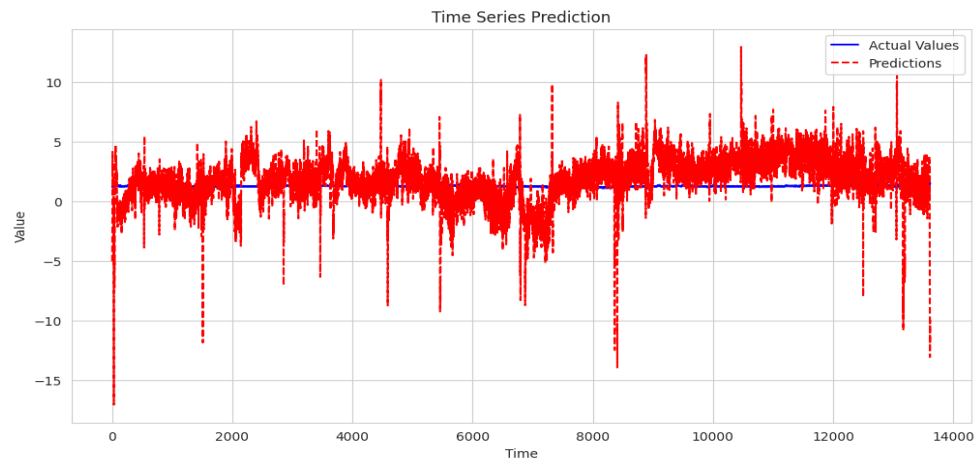
Contents

- 데이터
 - 데이터 이해
 - 실험 데이터 전처리
- 모델링
 - 머신러닝 모델 (Linear regression, Random forest)
 - 딥러닝 모델 (1D CNN, LSTM)
- 향후 계획

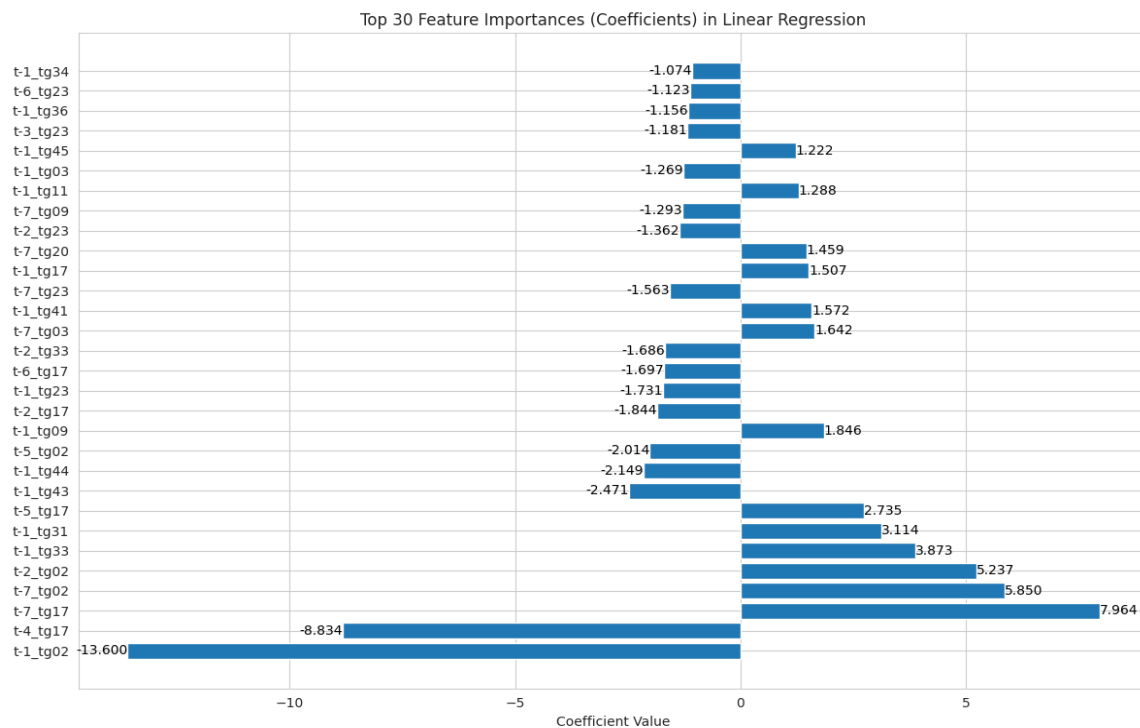
머신 러닝 모델 (Linear regression)

■ 예측 성능 지표

- R^2 : -1112.9390
- MSE : 4.0667



■ 변수 중요도 해석

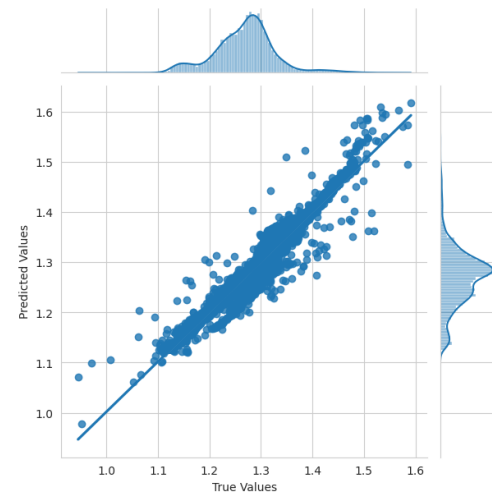
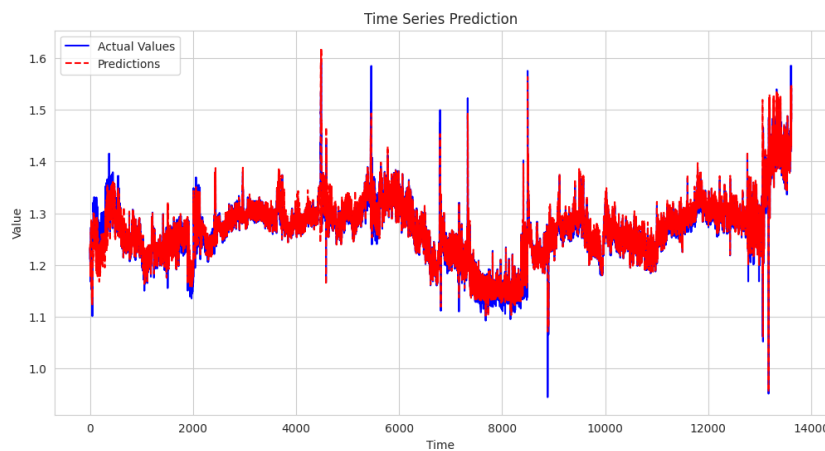


* 모델 예측 성능이 좋지 않아
변수 별 중요도를 신뢰하기 어려움

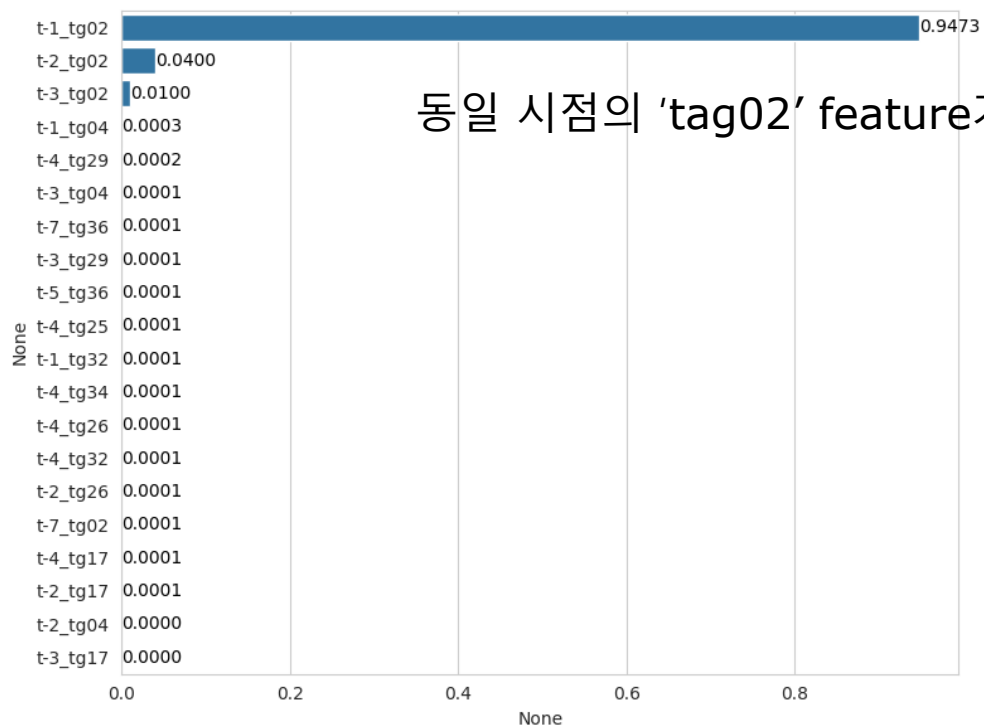
머신 러닝 모델 (Random forest)

■ 예측 성능 지표

- R^2 : 0.9644
- MSE : 0.0001301



■ 변수 중요도 해석



동일 시점의 'tag02' feature가 가장 중요하게 작용

딥러닝 모델 (1D CNN)

■ 모델 구조

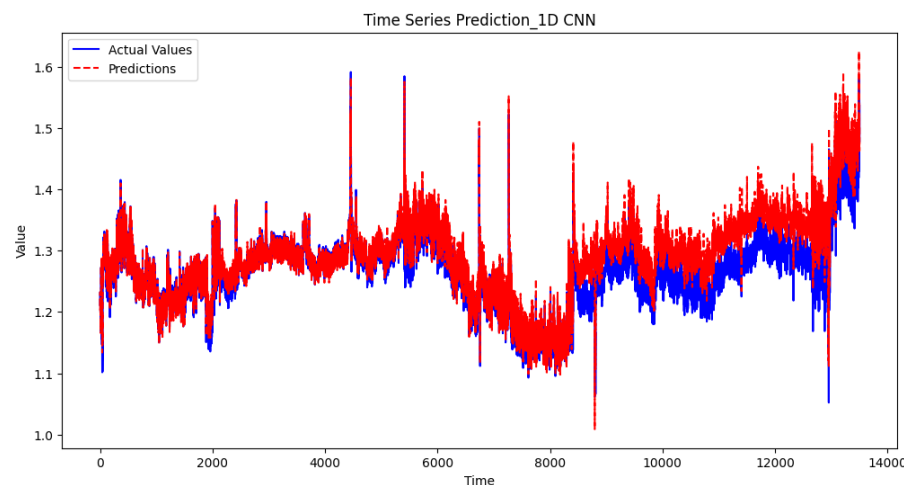
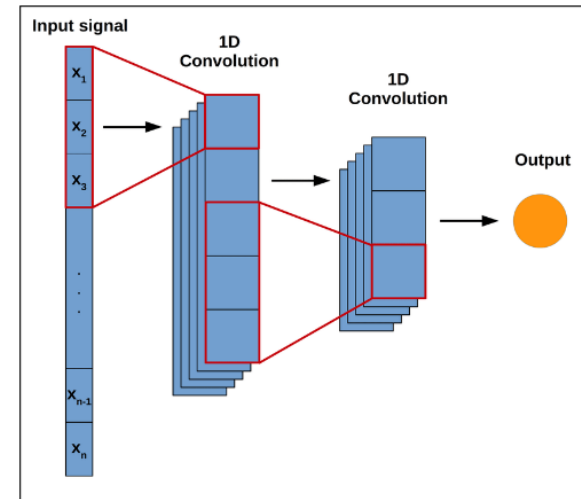
- 기존 CNN모델에서 input 및 kernel의 차원을 축소한 모델
- 시간에 따른 연속적인 패턴이나 특성을 식별하는데 유용
- 3개의 1D Conv layer(64-128-256 / kernel=3) + 3개의 linear layer (256-128-64-1)

■ Experiment setting

- Epoch : 200
- optimizer : Adam(lr=1e-4)

■ Result

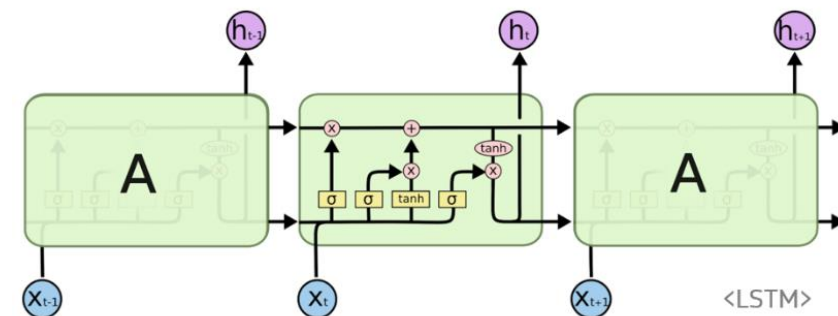
- R2 : 0.6970
- MSE : 0.0011
- MAE : 0.0242



딥러닝 모델 (LSTM)

■ 모델 구조

- RNN의 한 종류로 시퀀스 데이터 학습에 효과적
- LSTM layer(hidden=256, layer=6) + attention layer (모델 해석을 위함)

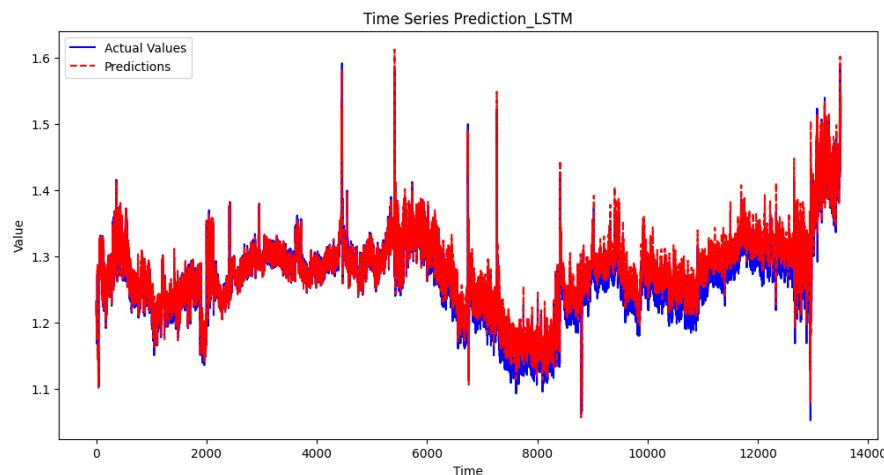


■ Experiment setting

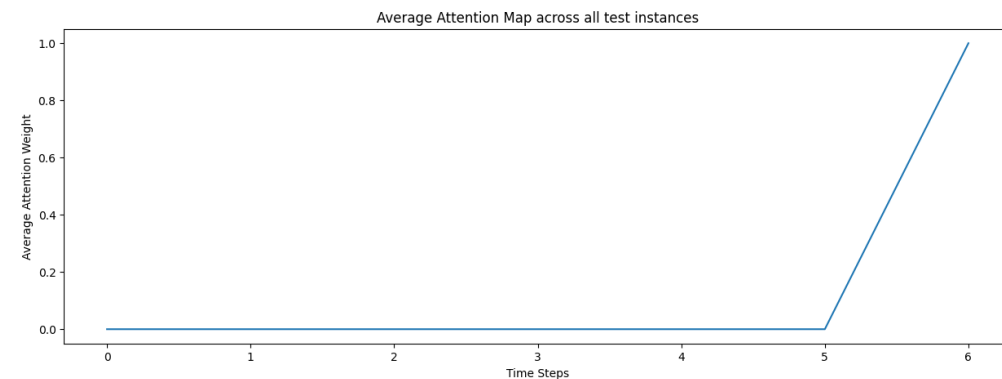
- Epoch : 200
- optimizer : Adam(lr=1e-4)

■ Result

- R2 : 0.9429
- MSE : 0.0002
- MAE : 0.0113

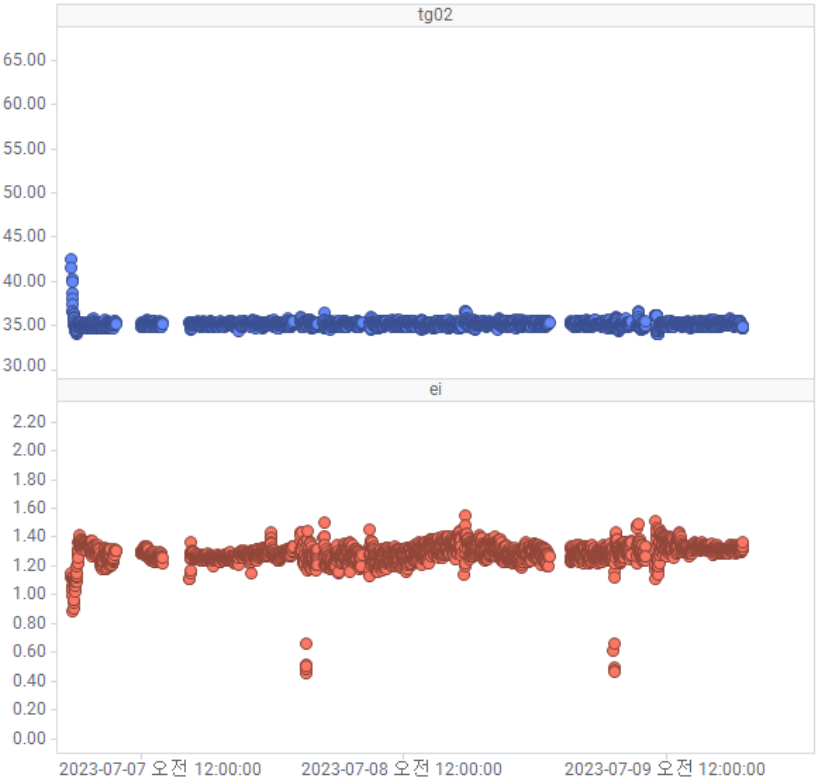
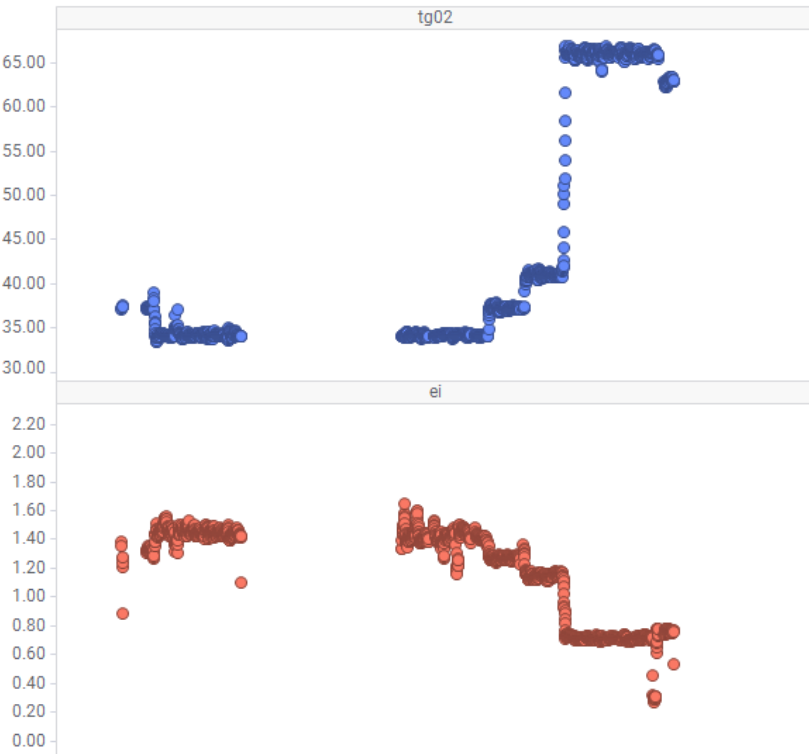
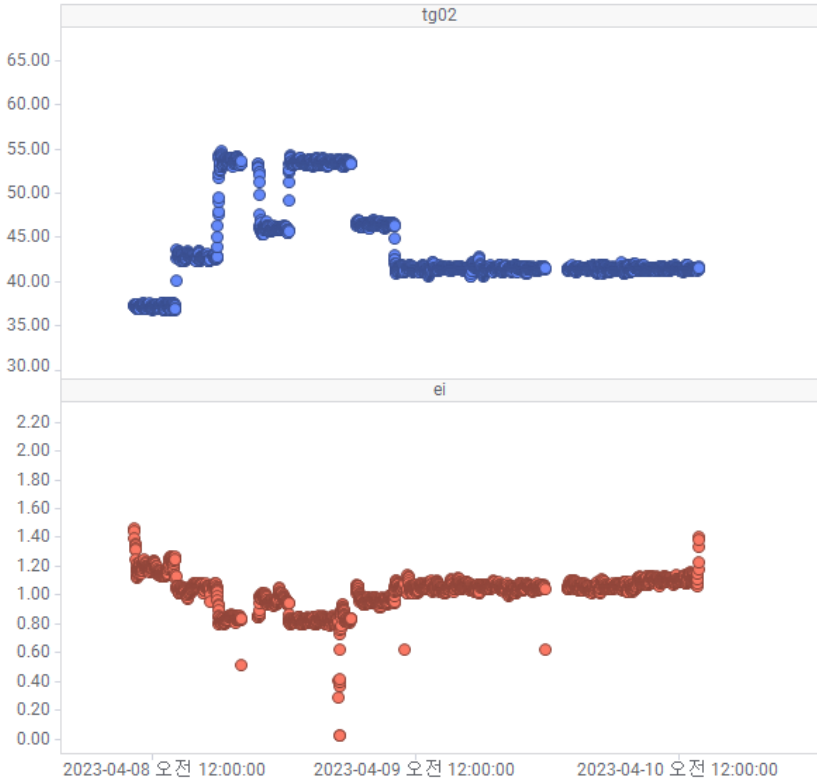


<Attention map>



과거 시점이 아닌 동일 시점의 영향력이 가장 큼

tg02 vs ei



Contents

- 데이터
 - 데이터 이해
 - 실험 데이터 전처리
- 모델링
 - 머신러닝 모델 (Linear regression, Random forest)
 - 딥러닝 모델 (1D CNN, LSTM)
- 향후 계획

향후 계획 및 논의

- 추가 방법론 구축/적용
 - 추가적인 딥러닝 모델 구축 – TCN (Temporal Convolutional Network) 등
 - 딥러닝 모델에 대한 해석 방법 개선 – Saliency Map, LRP, SHAP 등
- 동 시점을 제외한 과거 시점의 센서 데이터만을 활용한 모델링
- '이상'을 정의하는 방법
 - Classification을 활용한 접근 용이
- 의미 상 변수 선택
 - 각 센서의 의미 이해
 - 의미 상 중복 변수 제거 / 주요 변수를 선택

감사합니다