

23.12.14

다변량 스팀 사용 이상 감지 및 영향 변수의 원인 분석

서울과학기술대학교 데이터사이언스학과

이성호 sean0310@seoultech.ac.kr

배소희 shbae2819@g.seoultech.ac.kr

심재웅 jaewoong@seoultech.ac.kr

- Tg04 예측
- IMV-LSTM 활용
- 이상 탐지 정의 방안
- Tg04 vs 지절

데이터

- 이용 데이터

- 제품 2종에 대한 6달간 센서 데이터

`df_ext(2023-03,04)(5123,0385)_2023-11-16 seoultech`

`df_ext(2023-05,06)(5123,0385)_2023-11-16 seoultech`

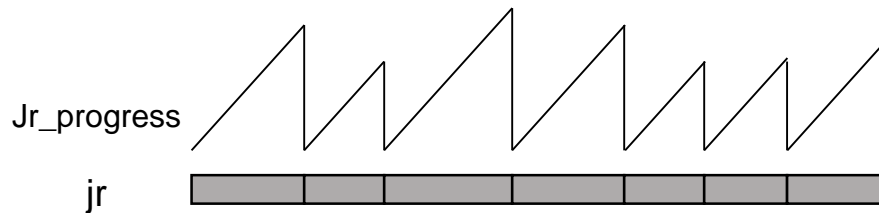
`df_ext(2023-07,08)(5123,0385)_2023-11-16 seoultech`

기간 : 2023-03-02 08:00:00 ~ 2023-08-27 03:00:00 (분)

date	날짜
tg	Sensor (38개)
stop	공정 분석값 0 : 가동 1 : 중지 이벤트 발생 2 : 중지 복구
jr	단위 공정값 / 제품 생산 주기 (생산품 번호)

- 이용 데이터 전처리

- jr_progress : jr을 기준으로 시간에 따른 정수를 새로운 변수로 추가
 - 전체 데이터에 대해 Minmaxscaler 적용



데이터

- 이용 데이터

- 이용 데이터 정의

- **input(X) : tg (37개 sensor 데이터), 공정진행도(jr_progress)**

- Window size : 20

- **output(y) : 10분 후의 tg 04(스팀 순간값)**

- Data split

- Train/Validation/Test = 60:20:20

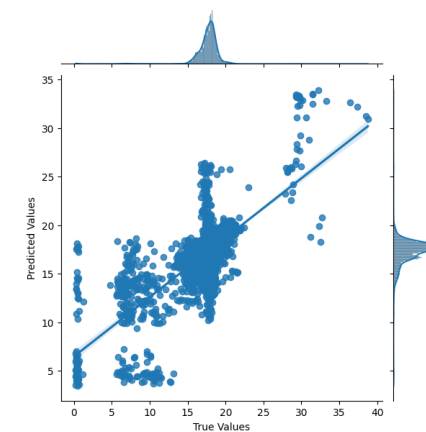
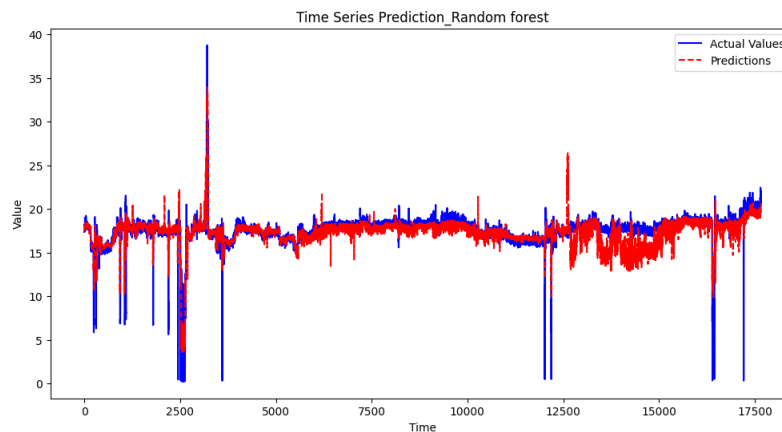
Train : 2023-03-02 08:00:00 ~ 2023-06-16 18:22:00 (52980)

Validation : ~ 2023-07-16 12:37:00 (17660)

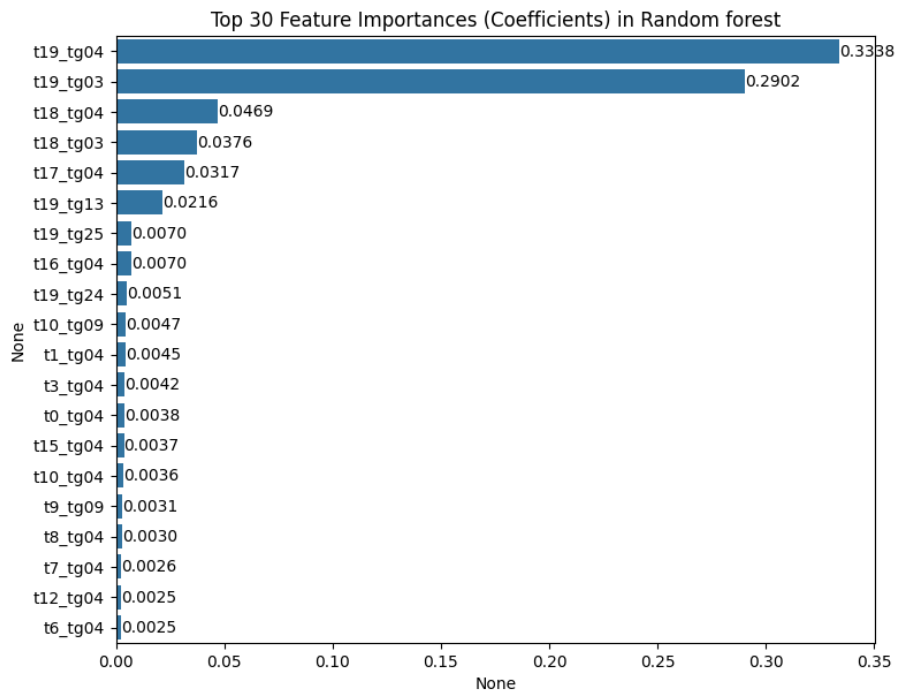
Test : ~ 2023-08-27 03:00:00 (17663)

머신 러닝 모델 (Random forest)

- 예측 성능 지표
 - R^2 : 0.4051
 - MSE : 2.2104



- 변수 중요도



tag04 & tg03가 가장 중요하게 작용

딥러닝 모델 (1D CNN)

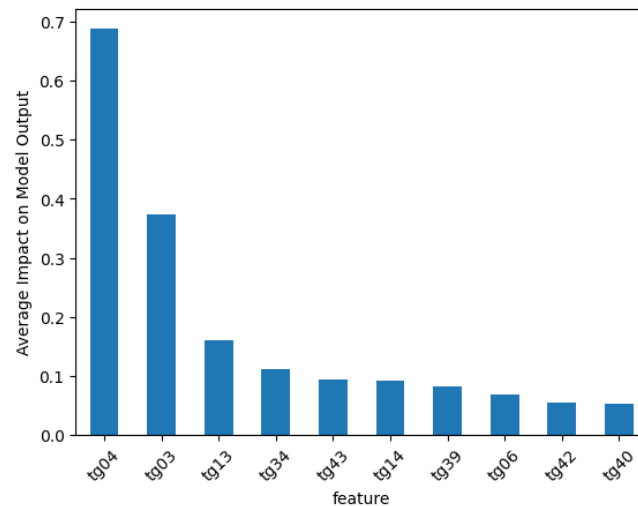
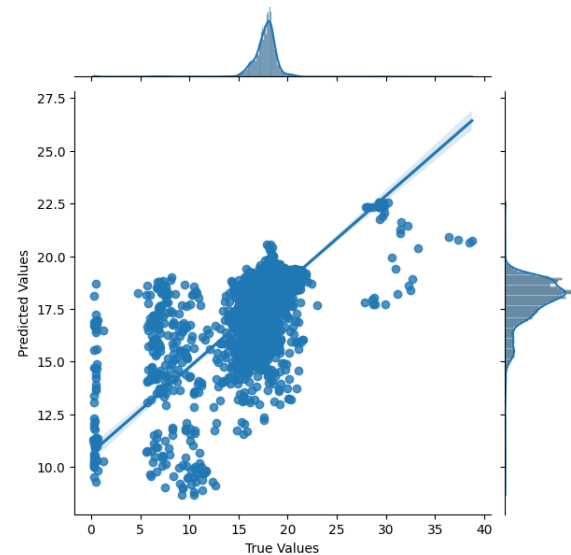
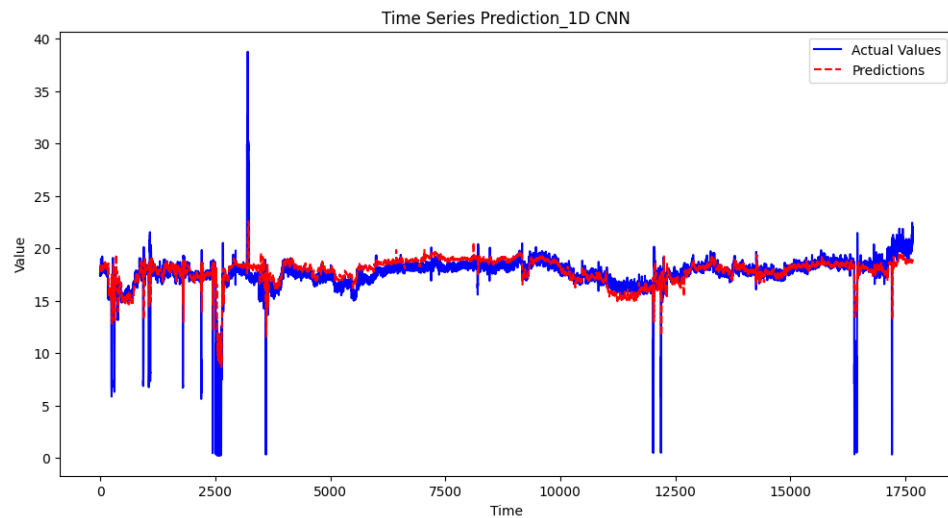
Experiment setting

- 1D Conv layers(64-128-256-512-1024 / kernel=3) + linear layers (1024-512-128-64-1)
- Epoch : 100
- optimizer : Adam(lr=1e-4)

Result

- R^2 : 0.4045
- MSE : 2.2125

SHAP을 통한 모델 해석



tg04: 스팀 순간값
tg03: 끝단 설비 속도
tg13: 설비B1 절대습도
tg34: 설비AE 절대습도
tg43: 설비 PE2 온도

딥러닝 모델 (LSTM)

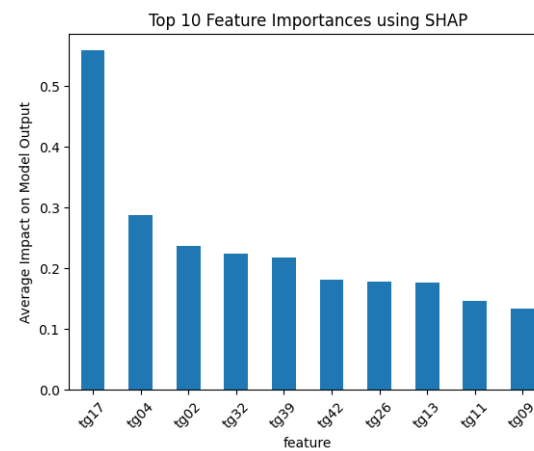
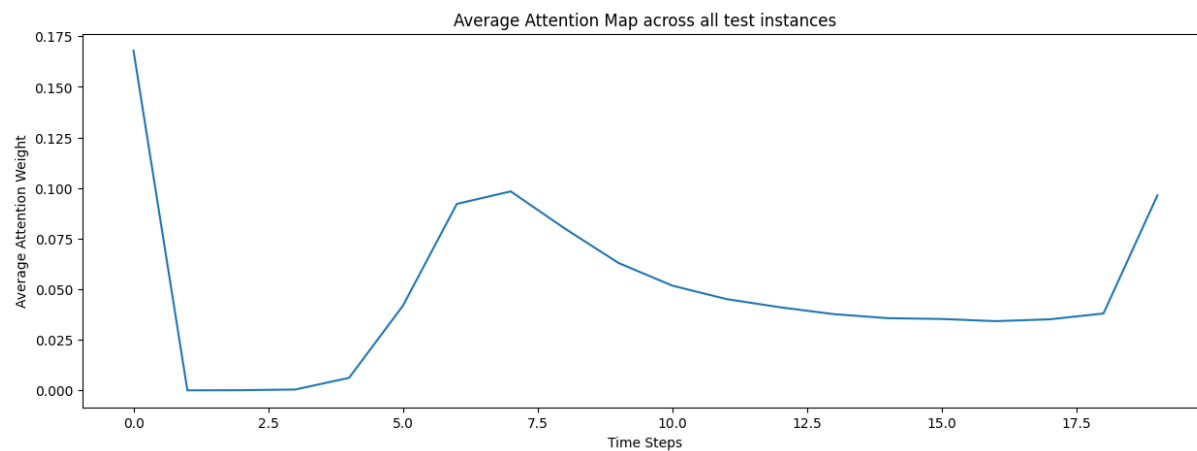
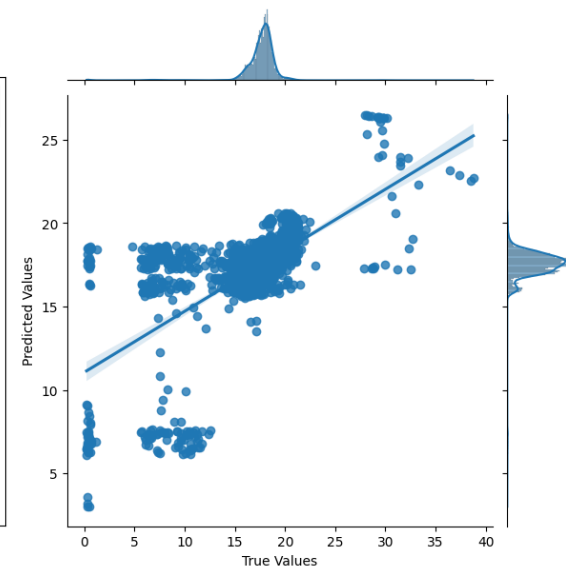
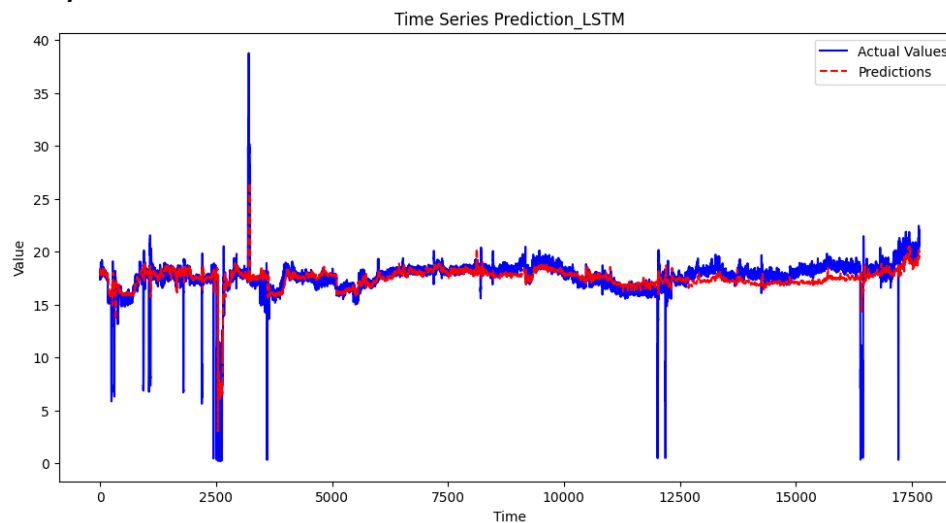
Experiment setting

- LSTM layer(hidden=256, layer=8) + attention layer
- Epoch : 100
- optimizer : Adam(lr=1e-4)

Result

- R^2 : 0.4018
- MSE : 2.2226

모델 해석



tg17: 스팀 누적값
tg04: 스팀 순간값
tg02: 종이별 측정 무게
tg32: 설비 S1 온도
tg39: 설비 AS 절대습도

딥러닝 모델 (DARNN)

- Experiment setting

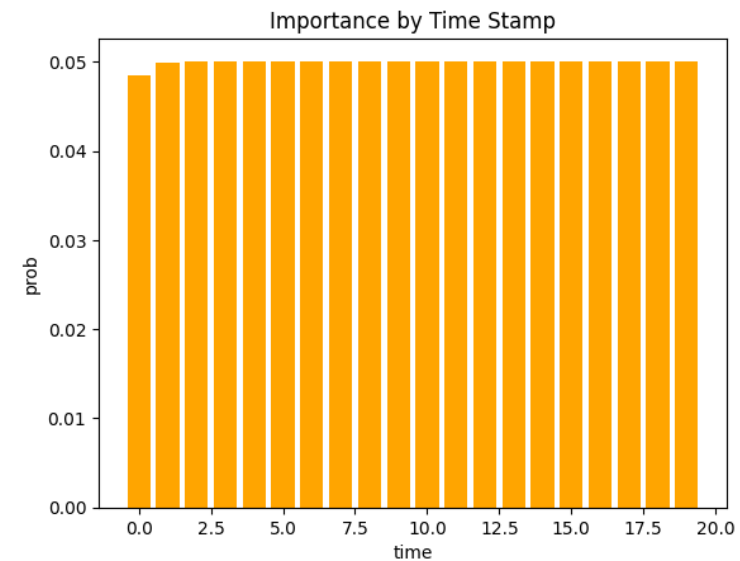
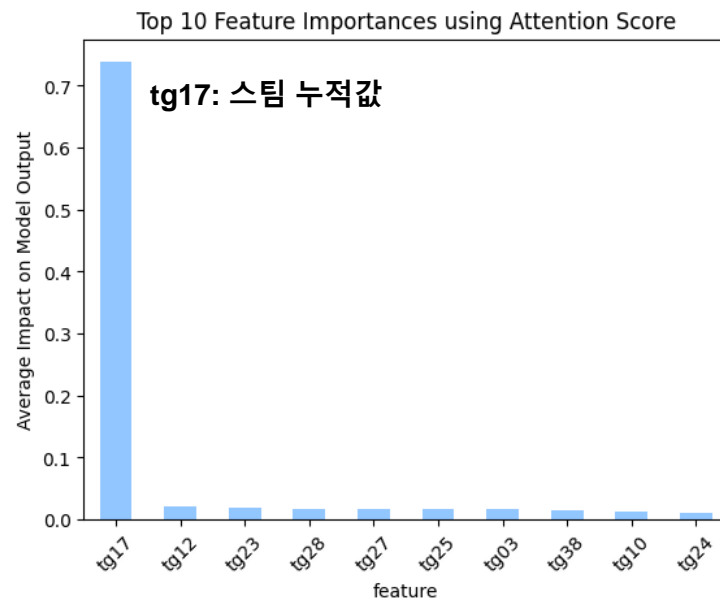
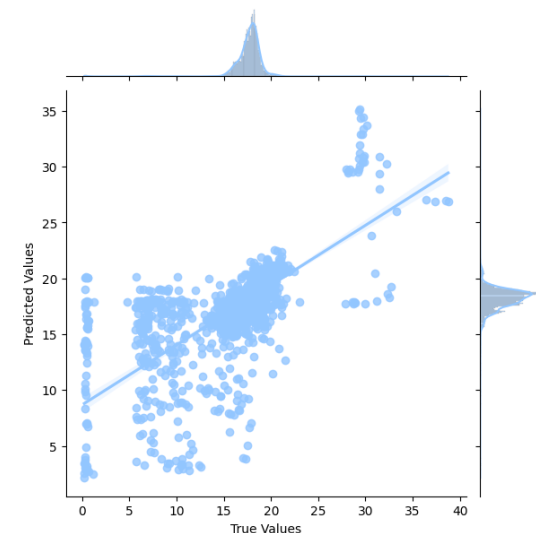
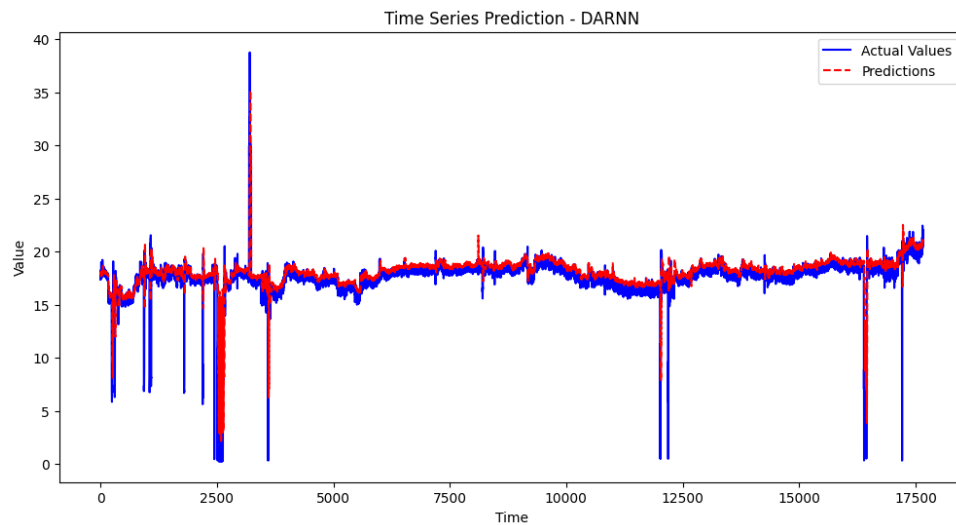
- Encoder & Decoder Layer : 16
- Epoch : 200
- Optimizer : Adam(lr=1e-3)

- Result

- R^2 : 0.4331
- MSE : 2.1064

- Attention Score를 통한 모델 해석

- tg17이 중요하게 작용



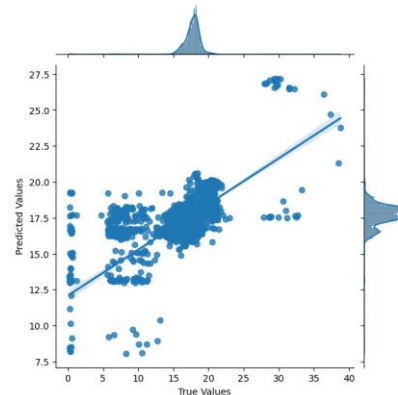
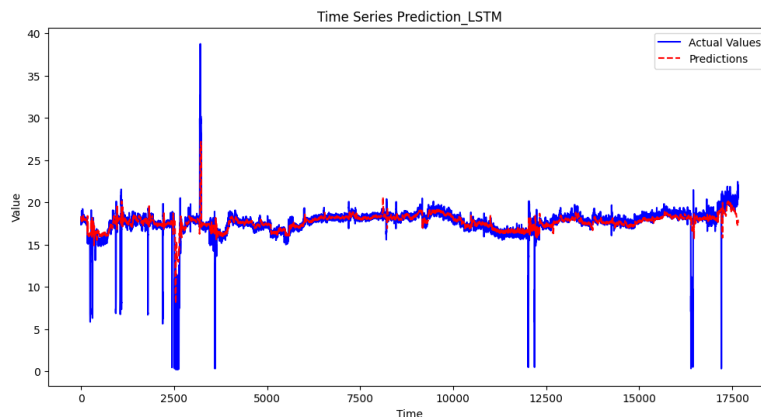
예측 지점과
먼 시점

예측 지점과
가까운 시점

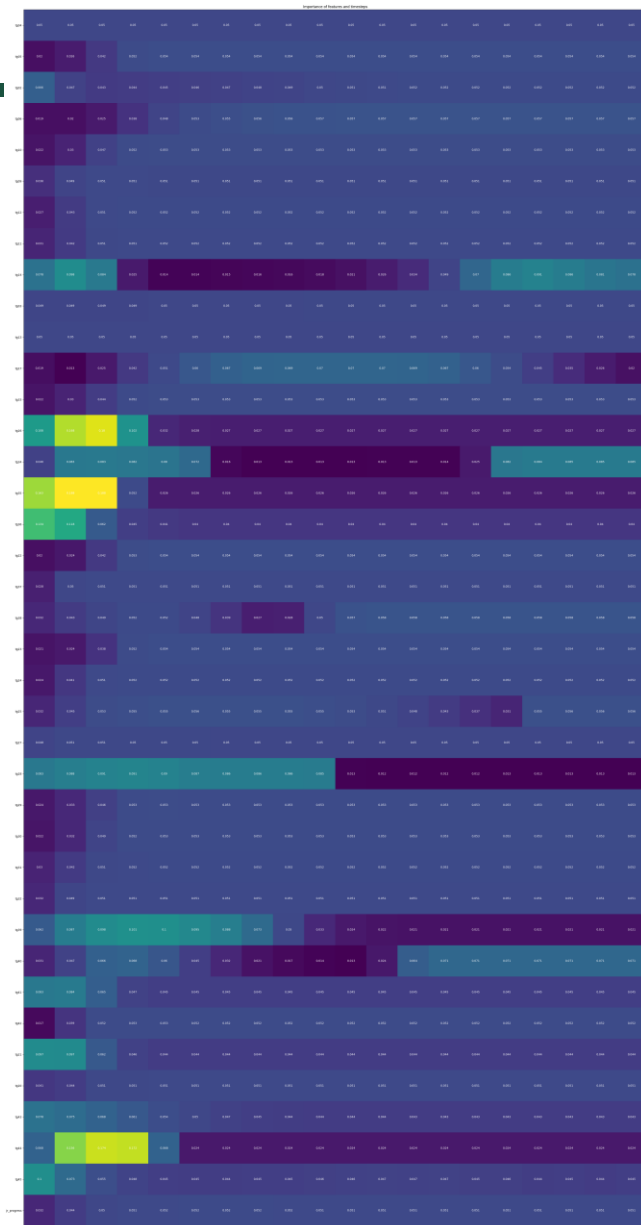
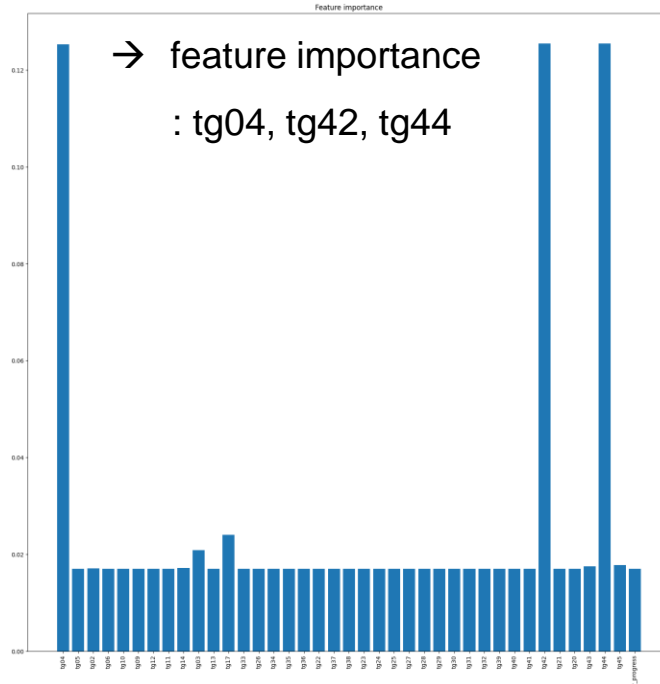
딥러닝 모델 (IMV-LSTM)

Guo, Tian, Tao Lin, and Nino Antulov-Fantulin. "Exploring interpretable LSTM neural networks over multi-variable data." *International conference on machine learning*. PMLR, 2019.

- Experiment setting
 - Lstm layer node : 32
 - Epoch : 100
 - Optimizer : Adam(lr=1e-3)
- Result
 - R^2 : 0.4136
 - MSE : 2.1787



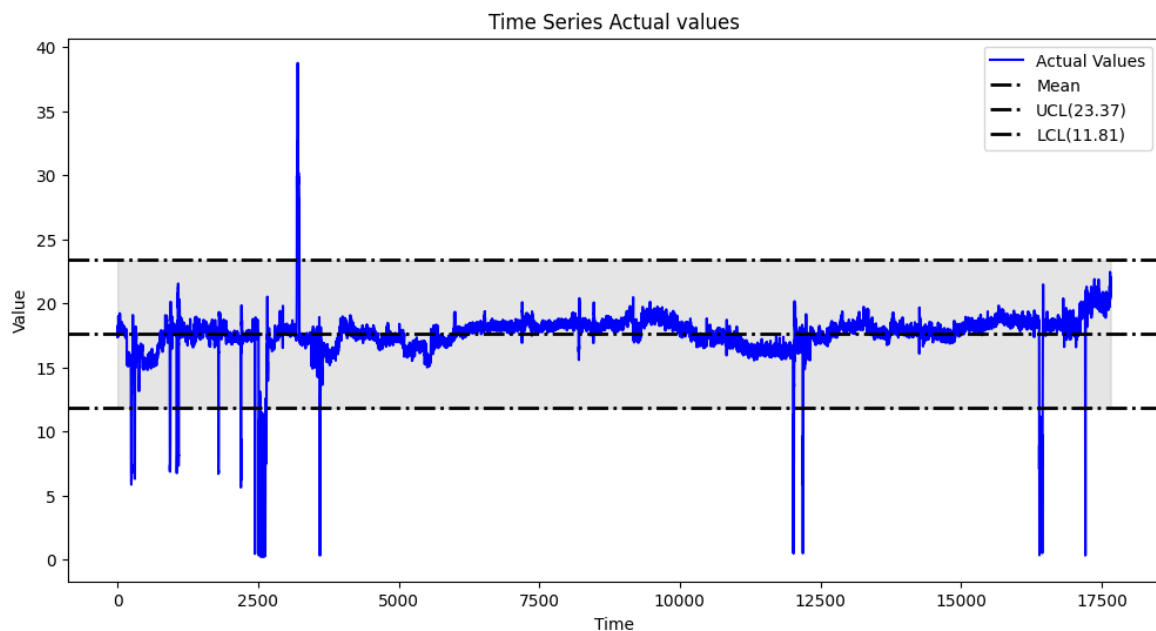
- Attention Map을 통한 모델 해석



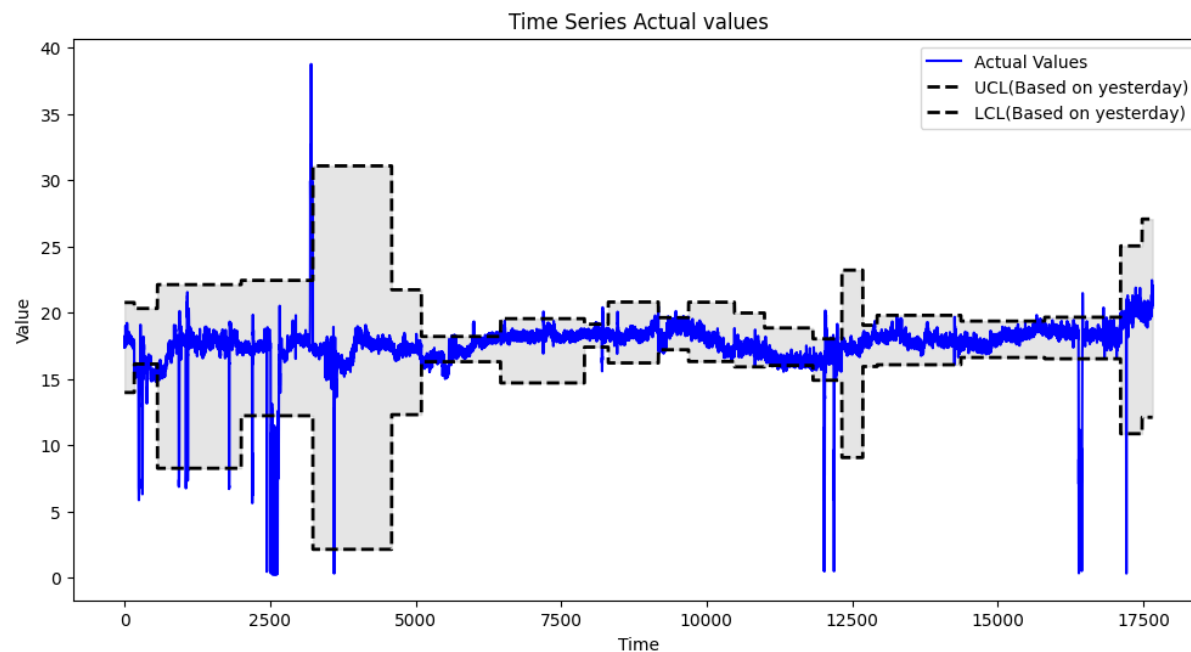
예측 지점과
먼 시점
예측 지점과
가까운 시점
<변수 별 temporal importance>

이상 임계값 설정 방안

- 전체 테스트 데이터셋 기준 $\text{mean} \pm 3\sigma$



- 하루 전 (60 * 24개 시점) 데이터셋 기준 $\text{mean} \pm 3\sigma$
- 지속 업데이트



Tg04 vs 지절



기타

- 단일 제품 데이터 혹은 제품명 태그 필요
 - 현재 데이터는 두 개 제품이 혼재되어 있음

감사합니다

이상 탐지

- Actual values
 - UCL/LCL(control limits) 기준의 시점과 stop 여부에 대해 확인
 - UCL/LCL은 전체 테스트 데이터셋으로 계산되며, Sigma 3기준

- Test구간
 - 2023-07-16 13:06:00 ~ 2023-08-27 03:00:00 (17663)

Test 내

stop 변수

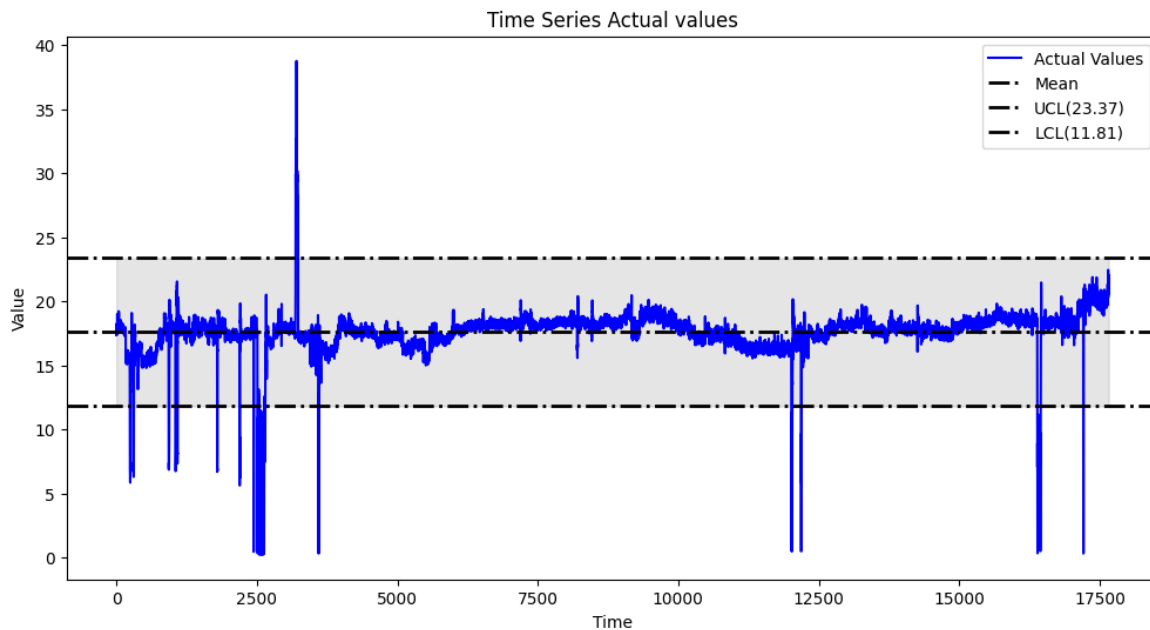
0	16134
1	654
2	875

- LCL이하, UCL이상인 경우인 index (350)

Index 내

stop

0	60
1	289
2	1



이상 탐지

- Actual values
 - UCL/LCL(control limits) 기준의 시점과 stop 여부에 대해 확인
 - UCL/LCL은 하루 전의 데이터셋으로 계산되며, Sigma 3기준

- Test구간
 - 2023-07-16 13:06:00 ~ 2023-08-27 03:00:00 (17663)

Test 내

stop

0	16134
1	654
2	875

- LCL이하, UCL이상인 경우인 index (1144)

Index 내

stop

0	726
1	325
2	93

