24.01.19

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

Tg04에 대한 예측

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

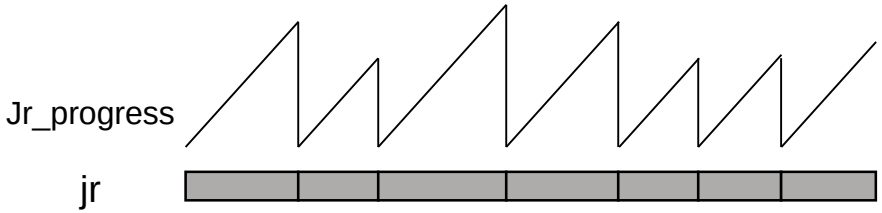
df_ext(2023-07,08)(5123,0385)_2023-11-16 seoultech

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

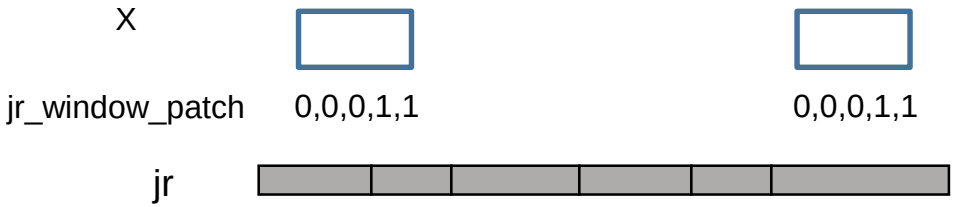
date	날짜
tg	Sensor (38개)
stop	공정 분석값 0 : 가동 1 : 중지 이벤트 발생 2 : 중지 복구
jr	단위 공정값 / 제품 생산 주기 (생산품 번호)

– 이용 데이터 전처리

• jr_progress : jr을 기준으로 시간에 따른 정수를 새로운 변수로 추가



• jr_window_patch : window 사이즈 만큼의 데이터 패치 후, 예측하고자 하는 시점의 jr과 다른지 같은지 표기하는 새로운 변수 추가



데이터

- 이용 데이터

– 이용 데이터 정의

- **input(X) : tg (38개 sensor 데이터), 공정진행도(jr_progress) , 공정변화도(jr_window_patch)**

  - Window size : 30

- **output(y) : 5분 후의 tg 04(스팀 순간값)**

– Data split

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

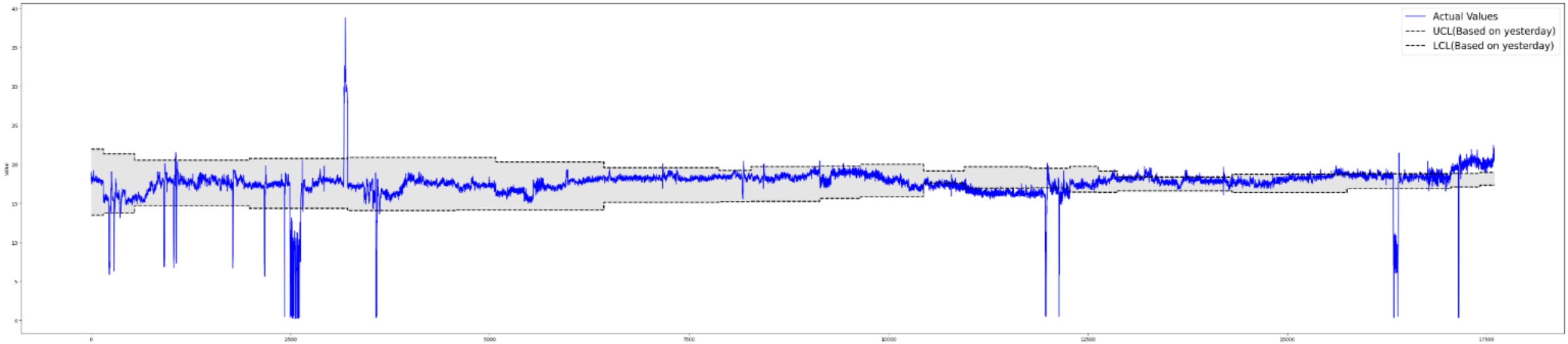
  Train : 2023-03-02 08:00:00 ~ 2023-06-16 18:37:00 (51840)

  Validation : ~ 2023-07-16 12:47:00 (17340)

  Test : ~ 2023-08-27 03:00:00 (17268)

UCL & LCL 설정(이상치 제외한 이전의 5일 누적)

Sigma = 3
누적 날짜 = 5
{0(LCL-UCL 구간 내): 14585, 1(LCL 이하): 1834, 2(UCL 이상): 1179}



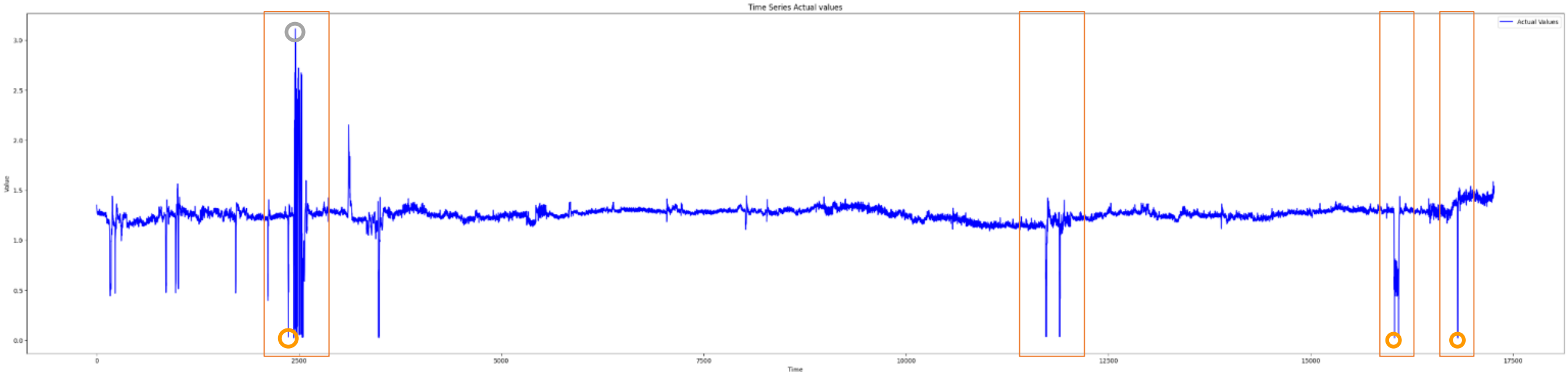
Instance 설정

- Value가 3이상인 index

index	date	ei
2452	2023-07-23 08:48:00	3.1125
2457	2023-07-23 08:53:00	3.0158

- Value가 0.025이하인 index

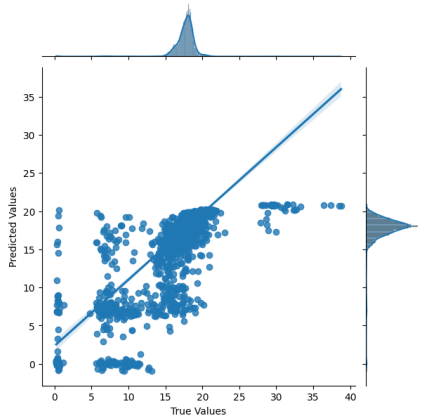
index	date	ei
2433	2023-07-23 08:29:00	0.0241
16036	2023-08-23 09:47:00	0.0241
16818	2023-08-26 19:31:00	0.0226



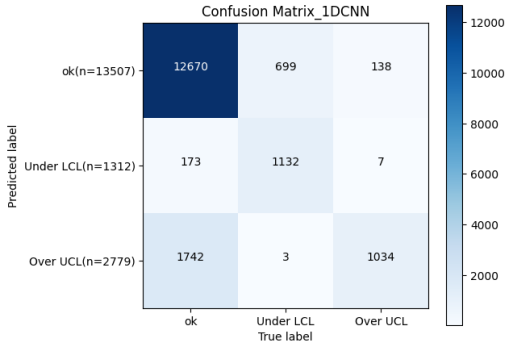
Industrial Artificial Intelligence Lab

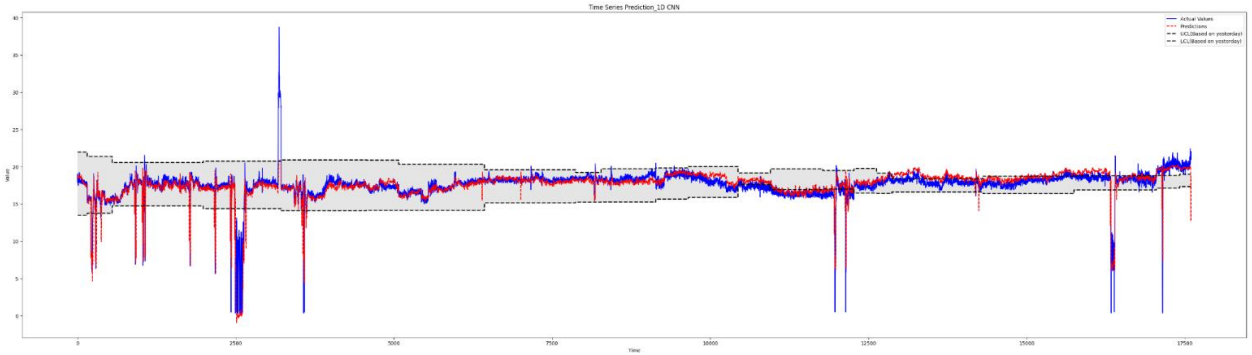
1D CNN

- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-4)
  - val loss가 가장 낮은 모델로 test
  - Layer : 1D Conv layer(16->32->64->128) / kernel=3),
    Linear layer (128 -> 32 -> 1)

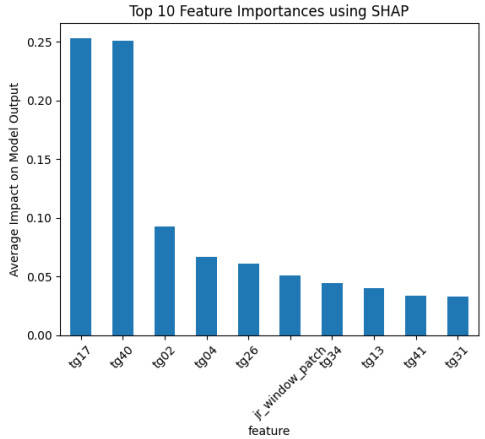


- Result
  - R^2 : 0.5151
  - MSE : 1.8069



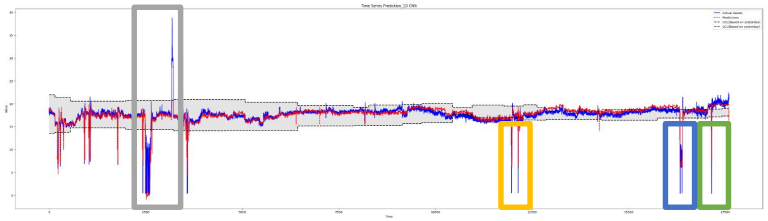


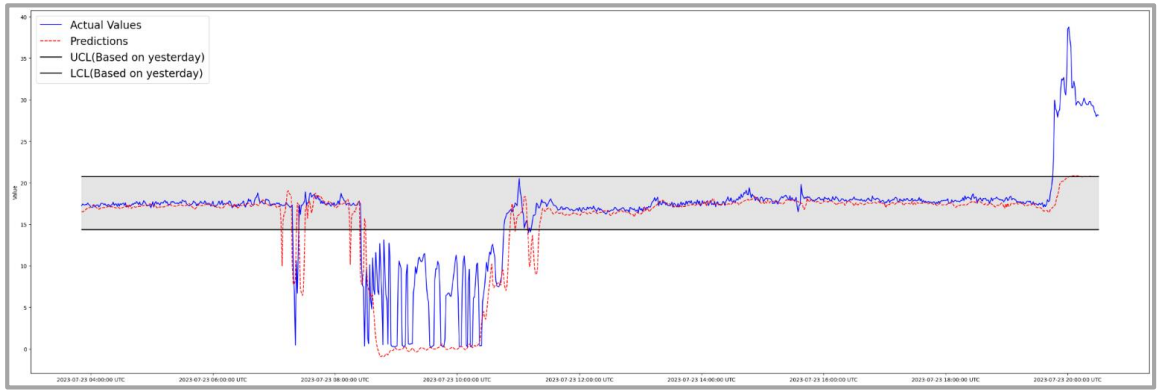
- 변수 별 중요도 해석
  - tg17, tg40

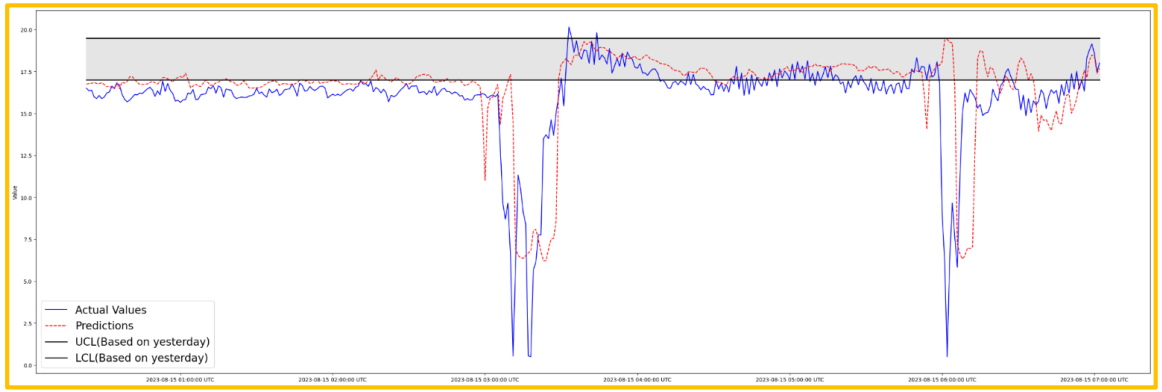


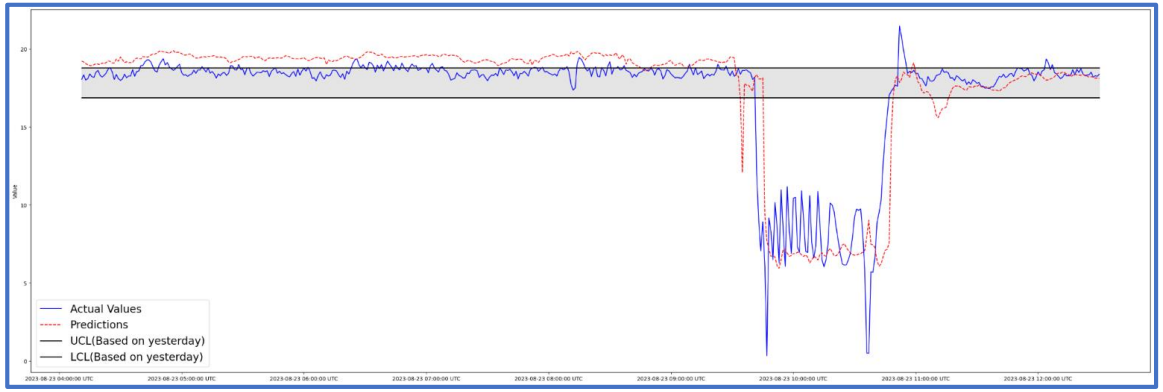
Industrial Artificial Intelligence Lab

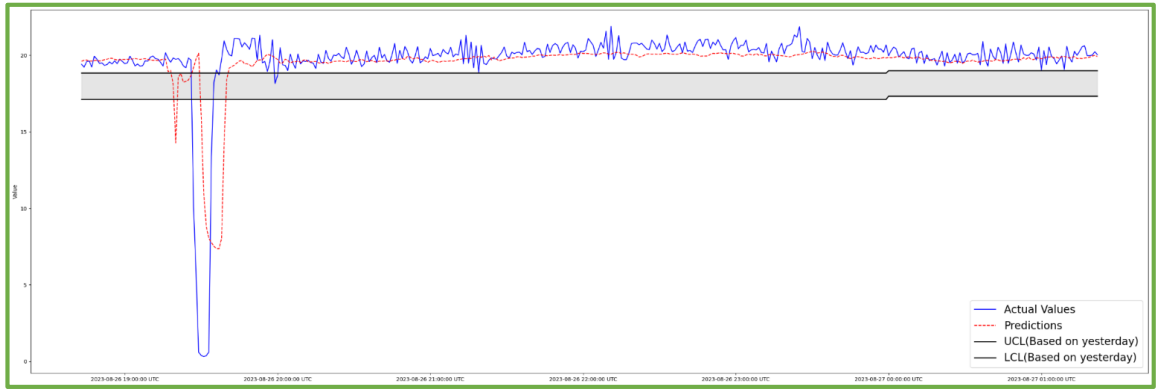
1D CNN







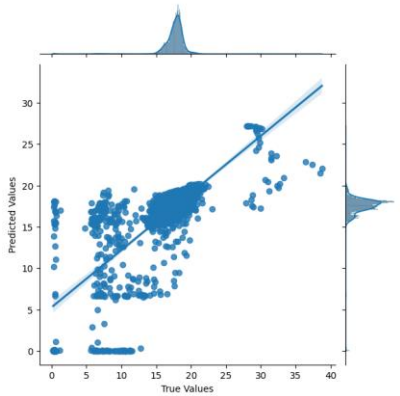


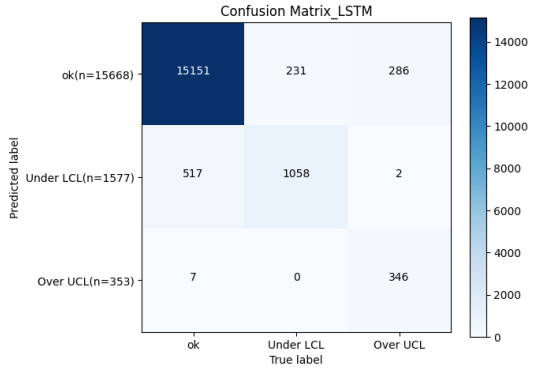


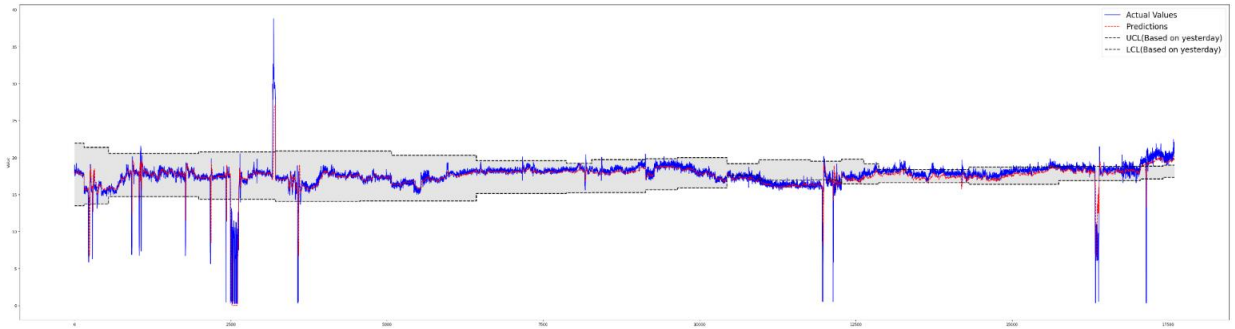
IAI
Industrial Artificial Intelligence Lab

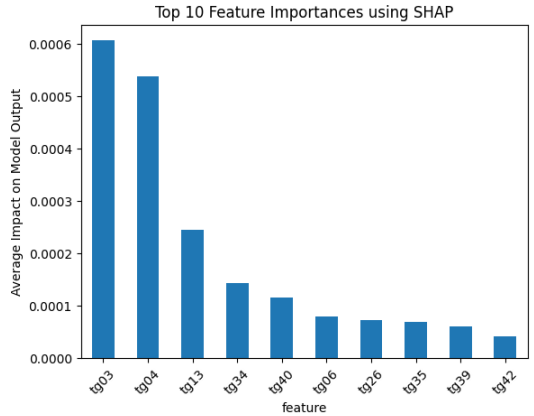
LSTM

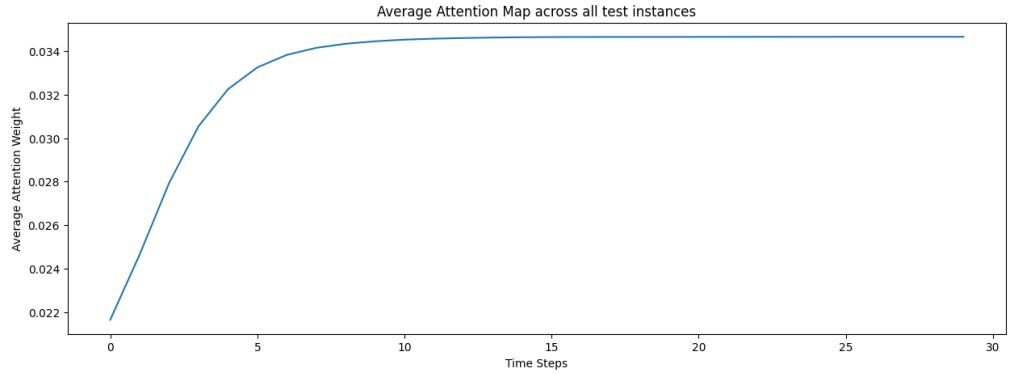
- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-4)
  - val loss가 가장 낮은 모델로 test
  - Layer : lstm layer(hidden=512, layer=6) + attention layer
- Result
  - R^2 : 0.5682
  - MSE : 1.6091
- 변수 별 중요도 해석
  - tg03, tg04





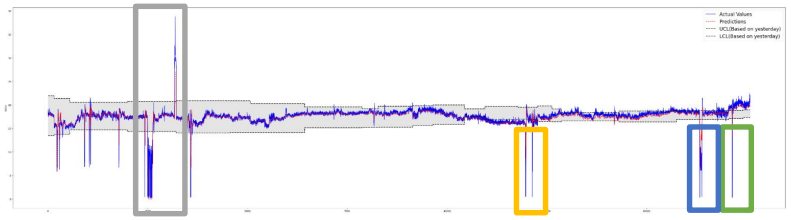


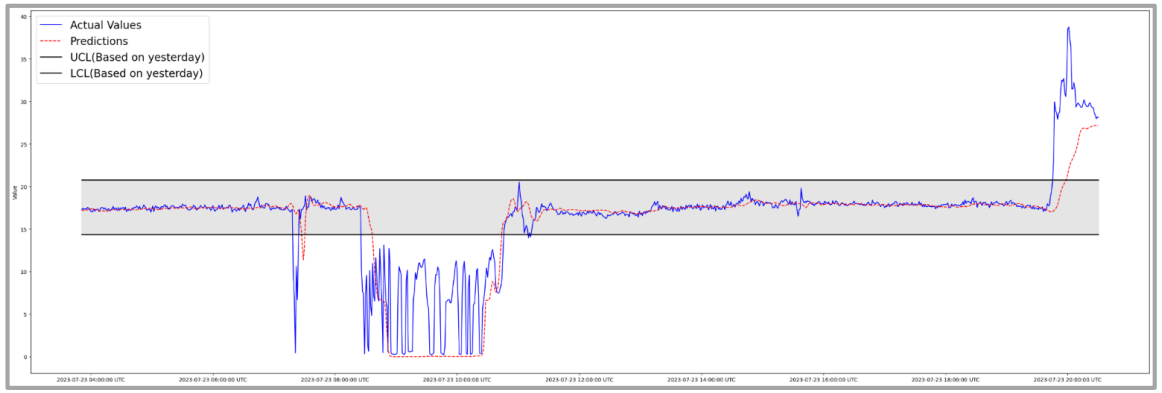


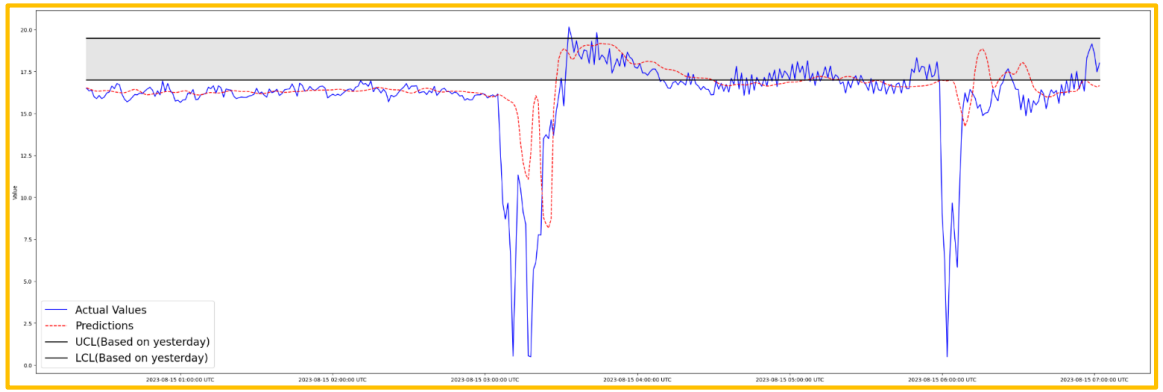


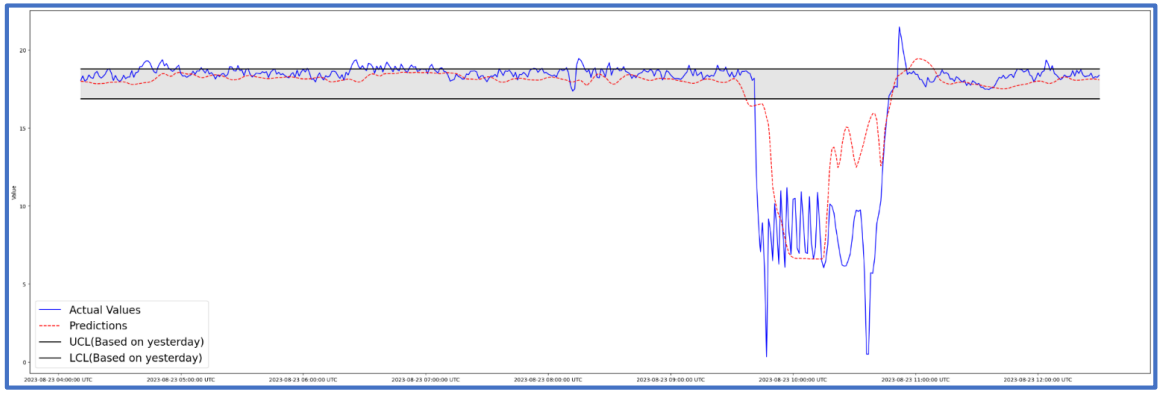
Industrial Artificial Intelligence Lab

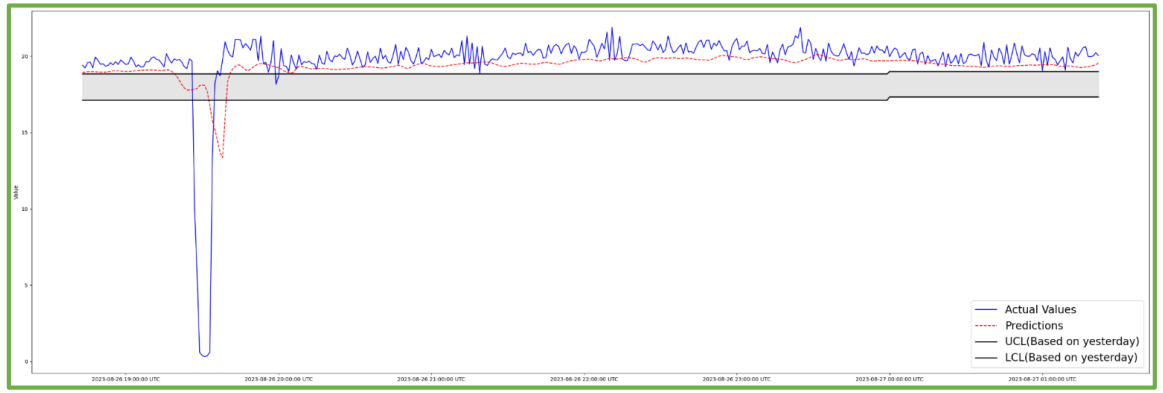
LSTM







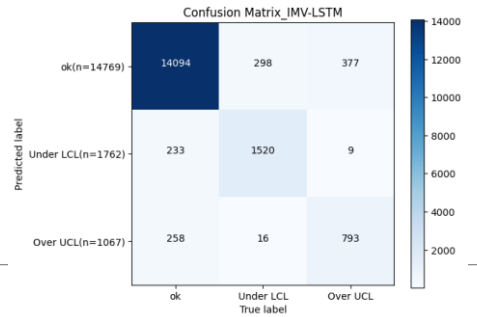


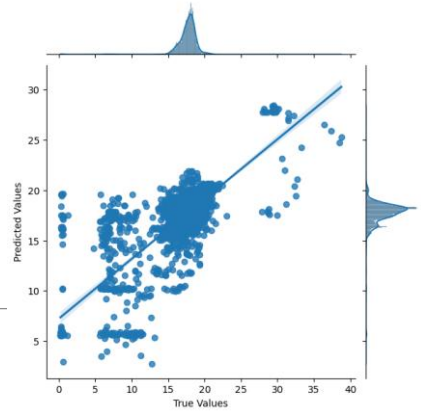


Industrial Artificial Intelligence Lab

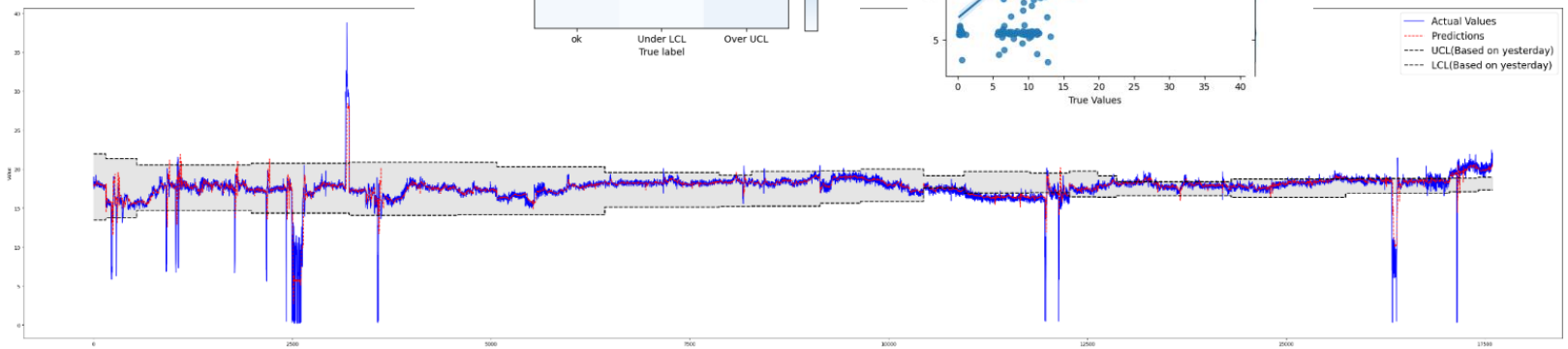
IMV-LSTM

- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-3)
  - val loss가 가장 낮은 모델로 test
  - Lstm layer 노드 : 32

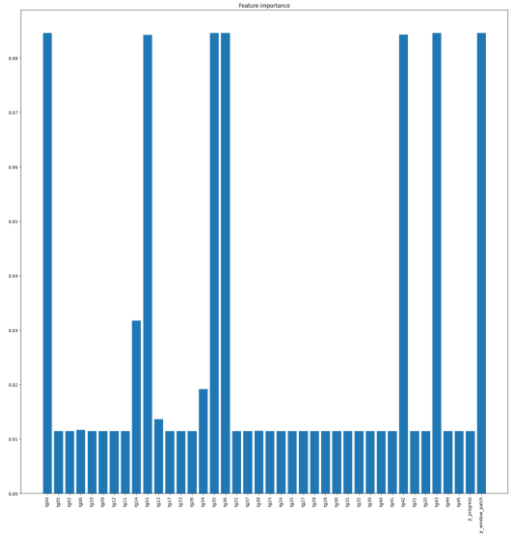




- Result
  - R^2 : 0.6028
  - MSE : 1.48
  - Acc : 0.9323
  - Macro F1 : 0.8373

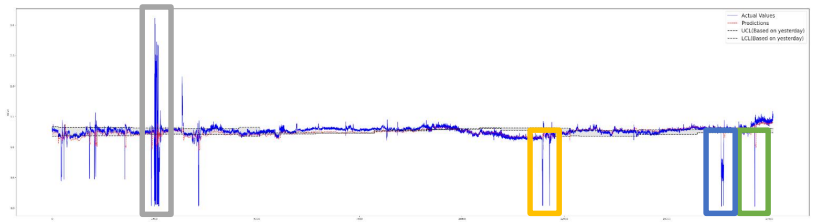


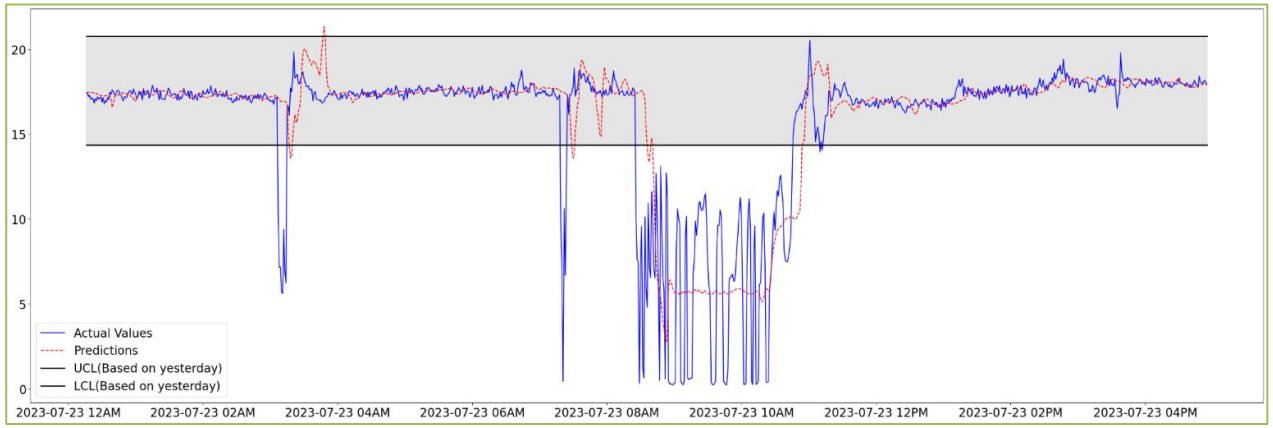
- 모델 해석(Attention Map)
  - tg04, tg03, tg35, tg36, tg42, tg43, jr_window_patch

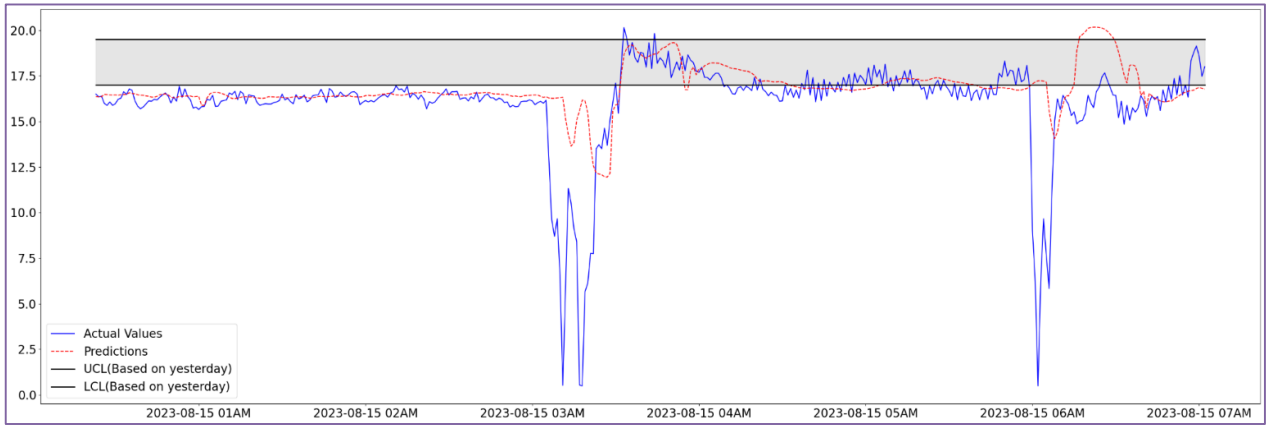


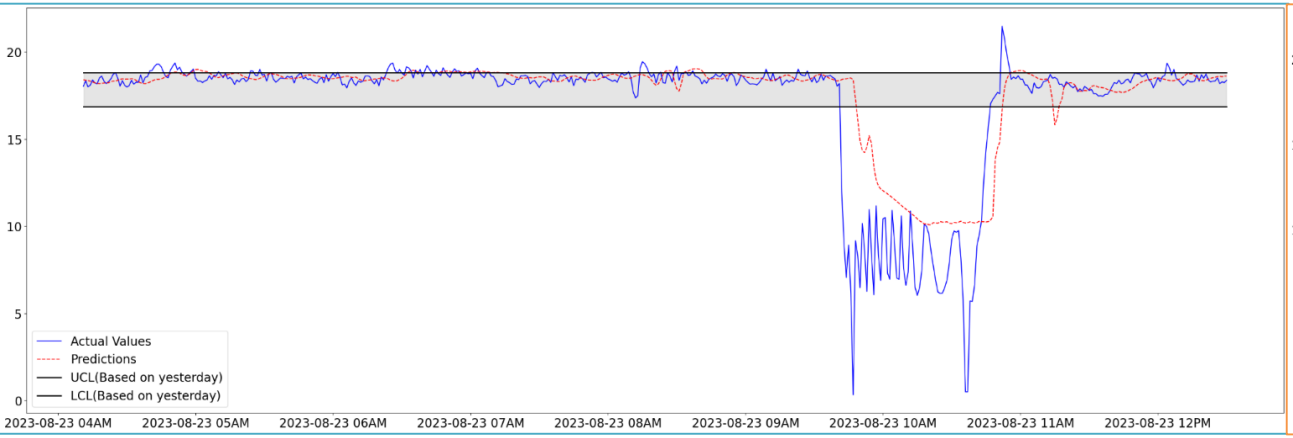
Industrial Artificial Intelligence Lab

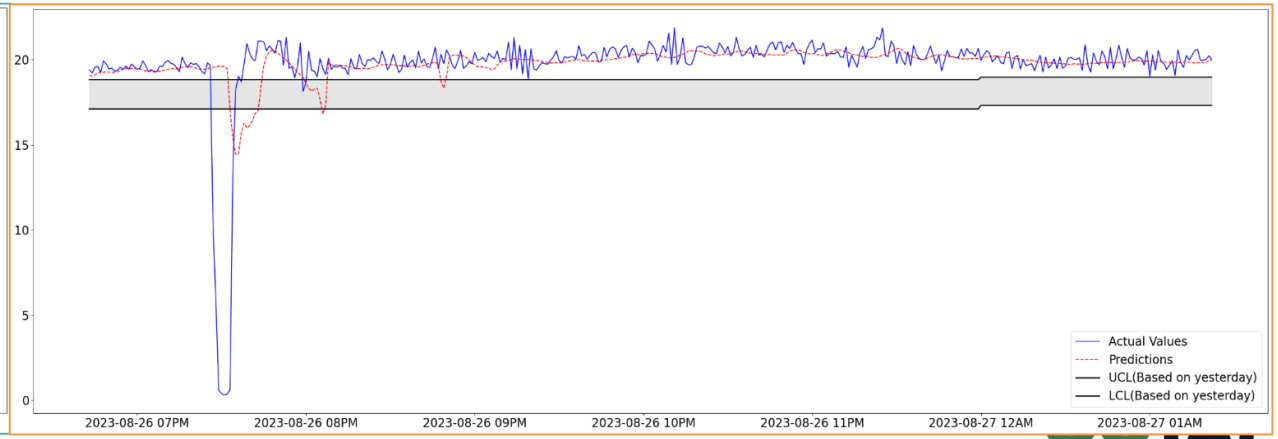
IMV-LSTM











Industrial Artificial Intelligence Lab

ei값에 대한 예측

데이터

- 이용 데이터

– 이용 데이터 정의

- **input(X) : tg (38개 sensor 데이터), 공정진행도(jr_progress) , 공정변화도(jr_window_patch)**

  - Window size : 30

- **output(y) : 5분 후의 ei계산값 tg04/(tg02*tg03*0.0003)**

– Data split

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

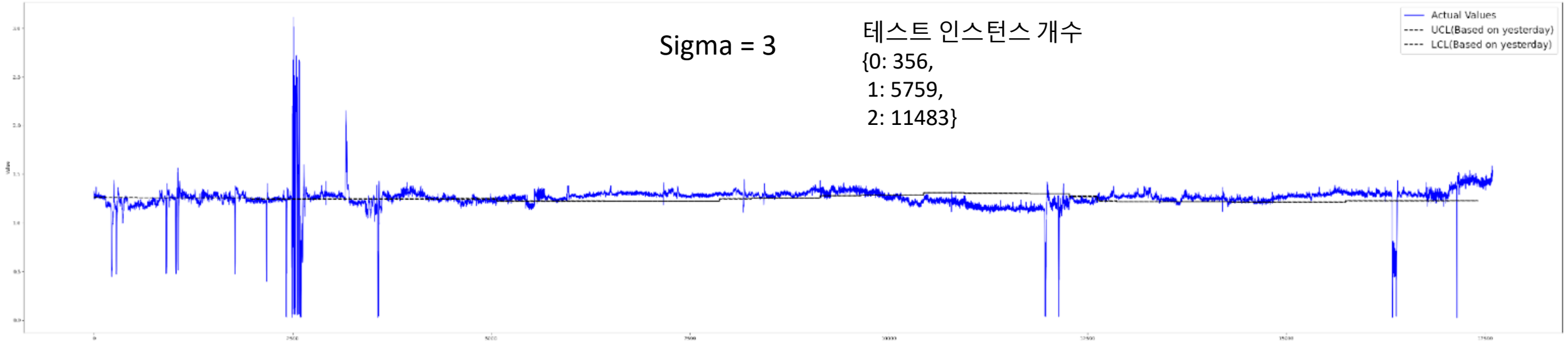
Train : 2023-03-02 08:00:00 ~ 2023-06-16 18:37:00 (51840)

Validation : ~ 2023-07-16 12:47:00 (17340)

Test : ~ 2023-08-27 03:00:00 (17268)

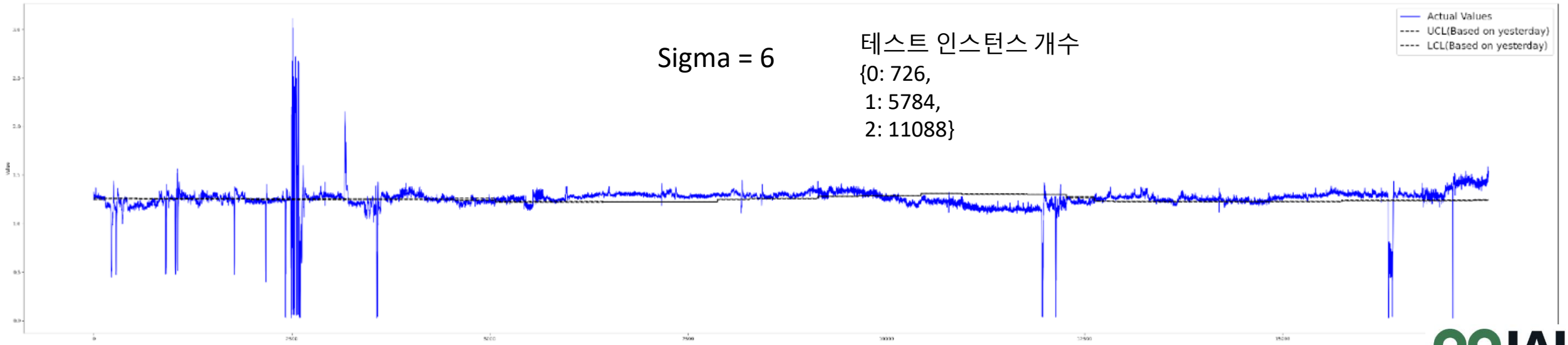
UCL & LCL 설정(이상치 제외한 이전의 7일 누적)

0 : LCL-UCL 구간 내
1 : LCL 이하
2 : UCL 이상



Sigma = 3

테스트 인스턴스 개수
{0: 356,
1: 5759,
2: 11483}



Sigma = 6

테스트 인스턴스 개수
{0: 726,
1: 5784,
2: 11088}

분산이 너무 작아 UCL – LCL 사이 구간이 너무 작아짐

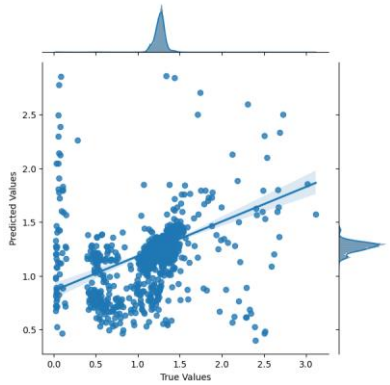
Industrial Artificial Intelligence Lab

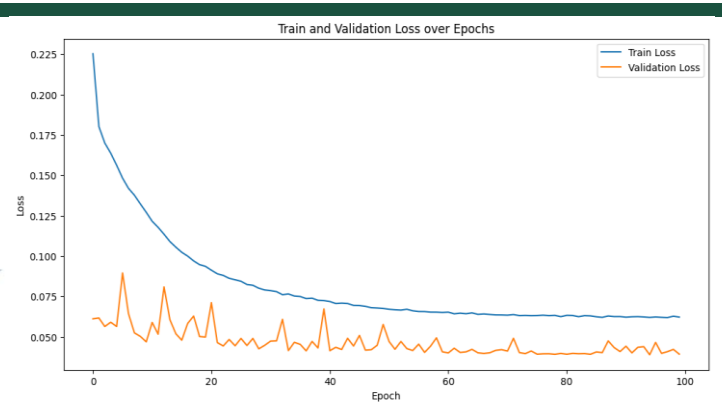
1D CNN

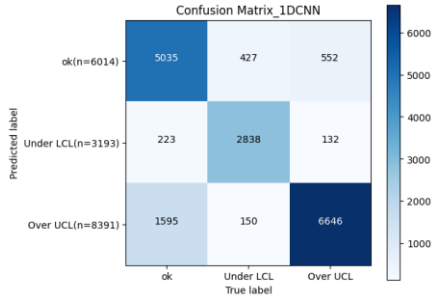
- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-4)
  - val loss가 가장 낮은 모델로 test
  - Layer : 1D Conv layer(16->32->64->128->256) / kernel=3),
    Linear layer (128 -> 32 -> 1)

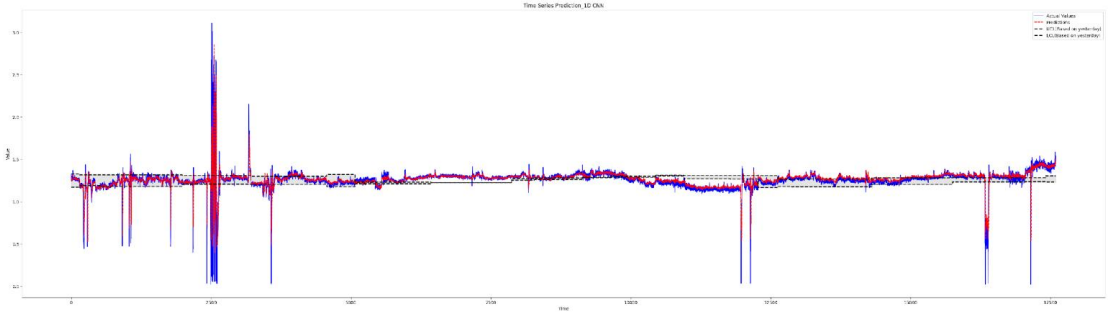
- Result
  - R^2 : 0.1317
  - MSE : 0.0146

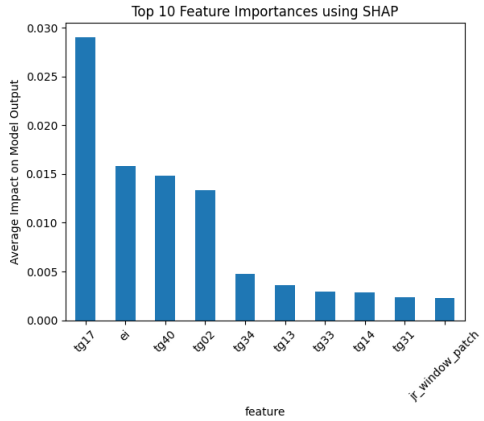
- 변수 별 중요도 해석
  - tg17, ei





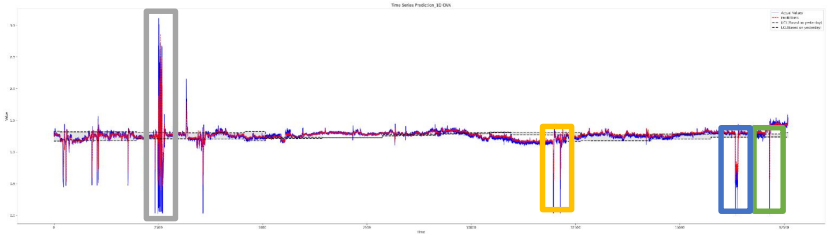


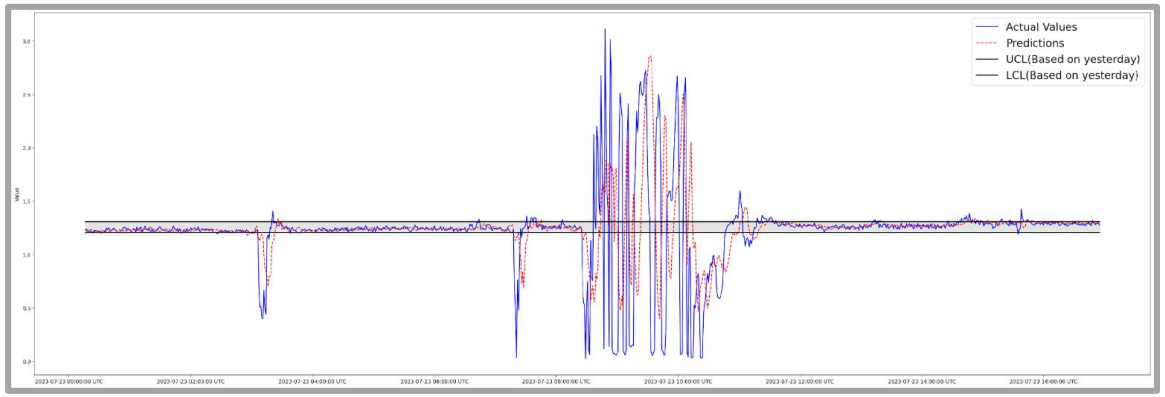


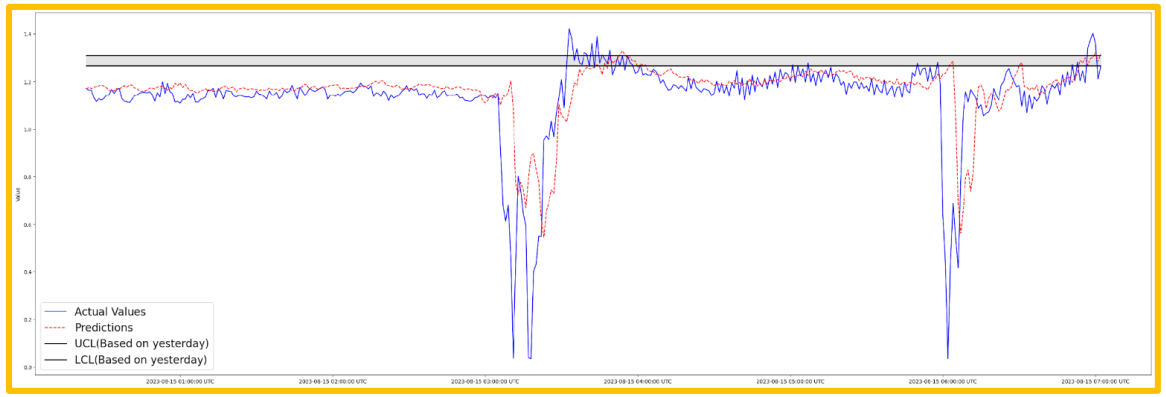


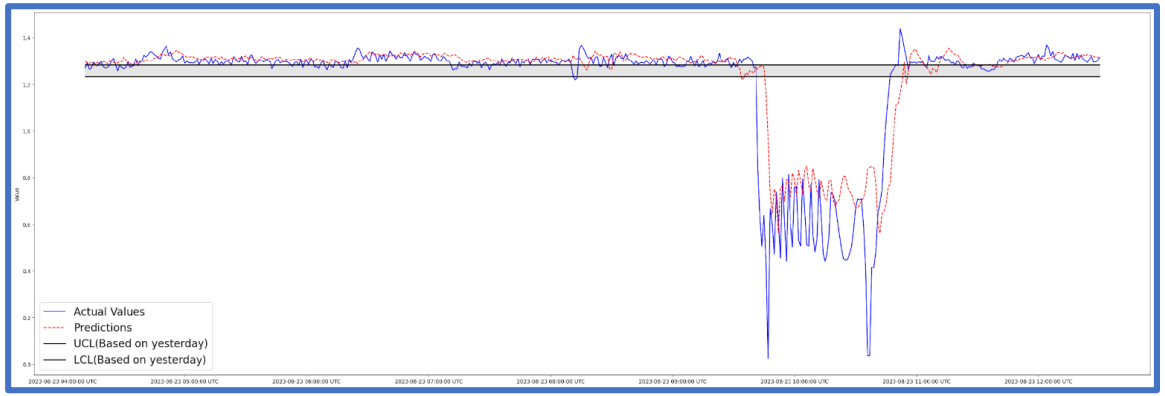
Industrial Artificial Intelligence Lab

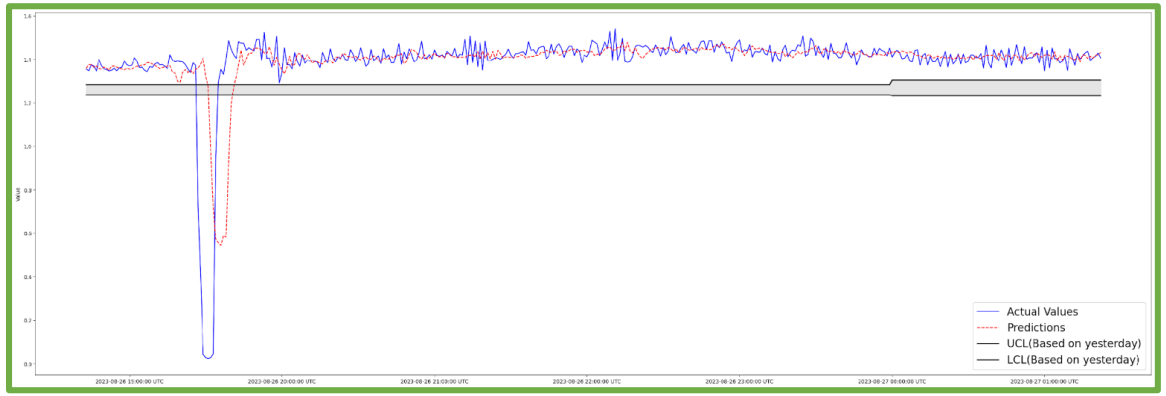
1D CNN







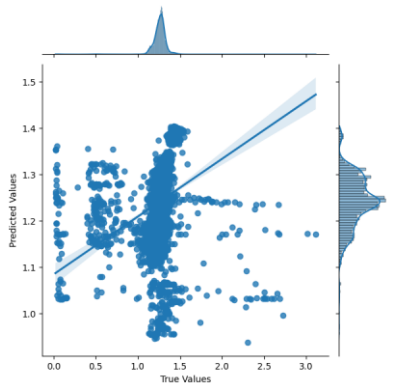


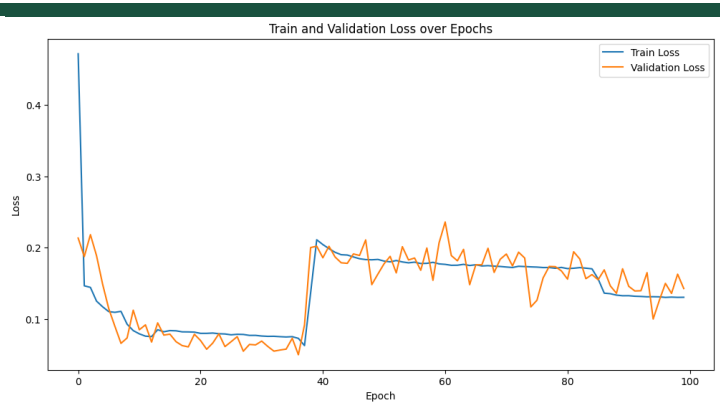


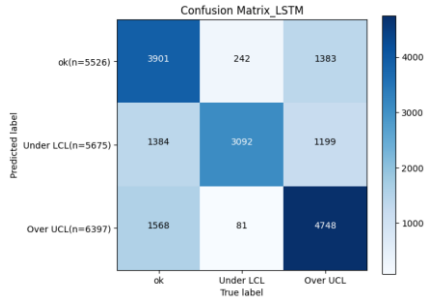
Industrial Artificial Intelligence Lab

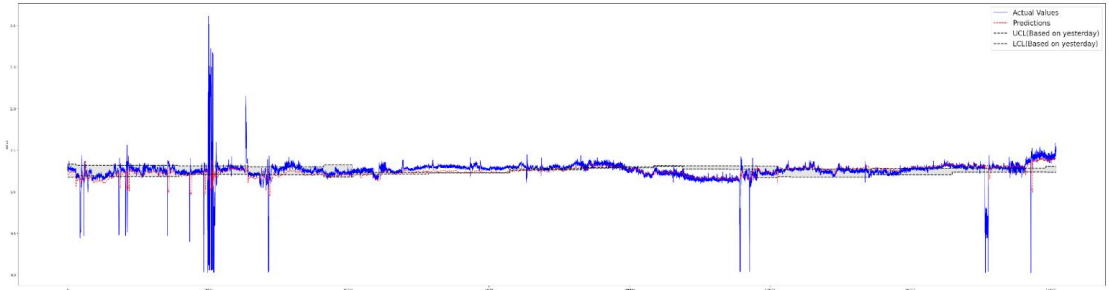
LSTM

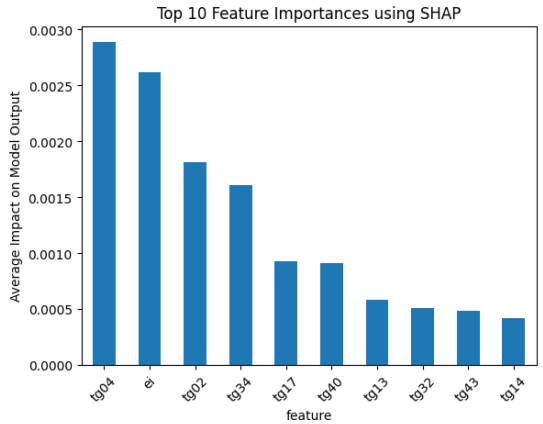
- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-4)
  - val loss가 가장 낮은 모델로 test
  - Layer : lstm layer(hidden=512, layer=6) + attention layer
- Result
  - R^2 : 0.0238
  - MSE : 0.0164
- 변수 별 중요도 해석
  - tg04, ei

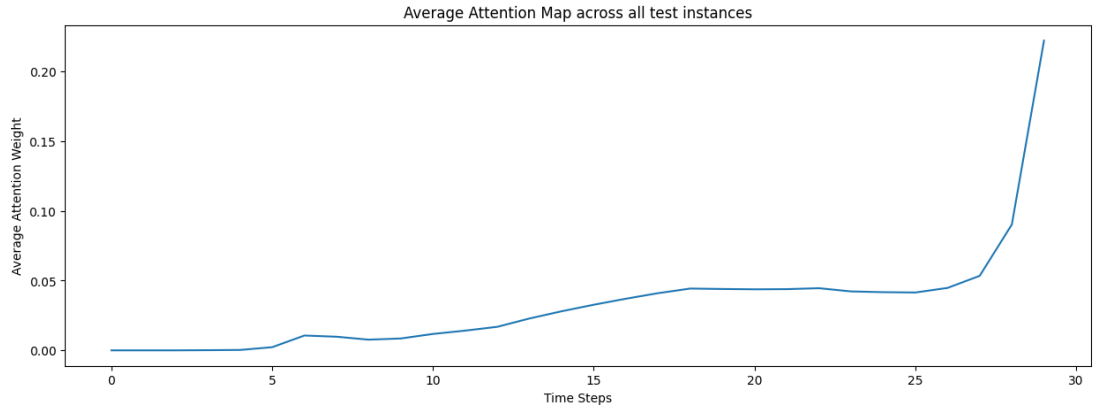








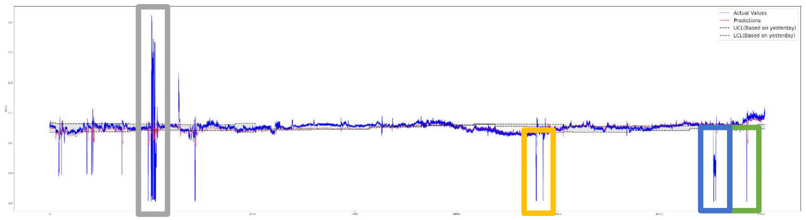


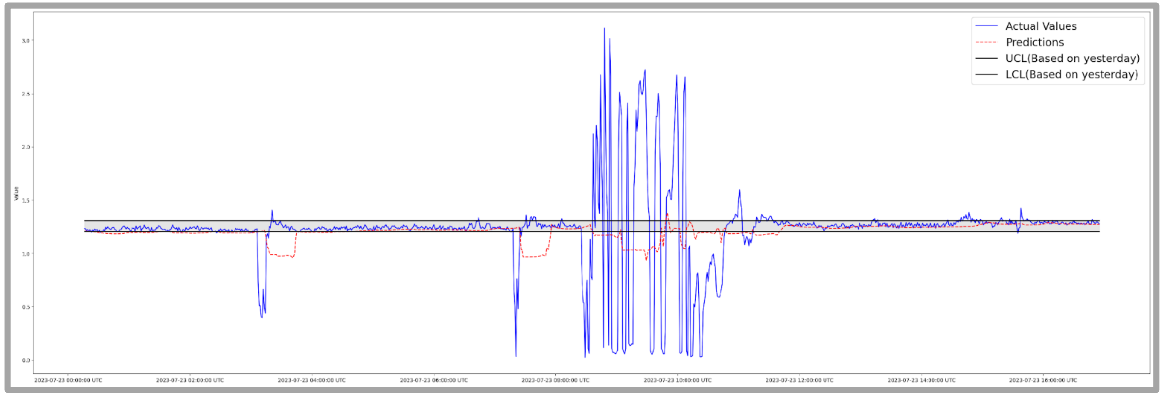


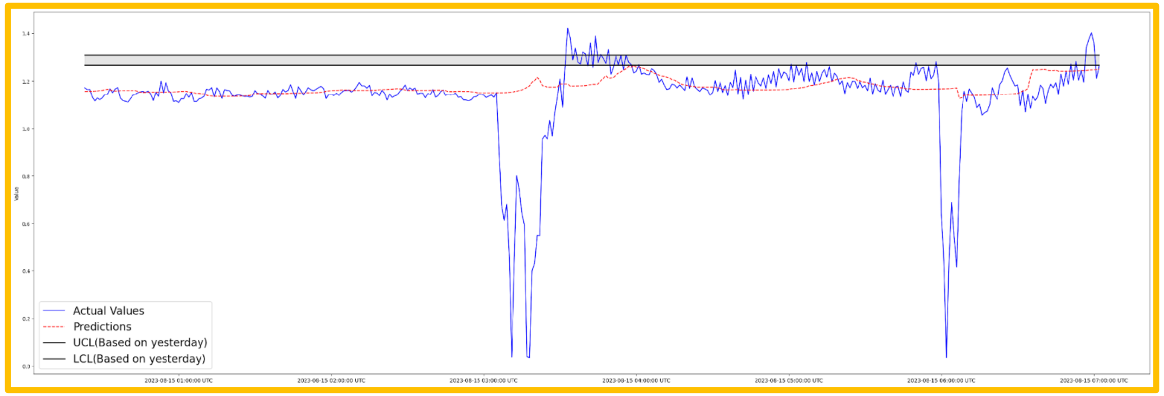
* Validation loss가 매우 불안정, epoch 조절해도 동일
* 예측값의 변동 폭이 매우 작음

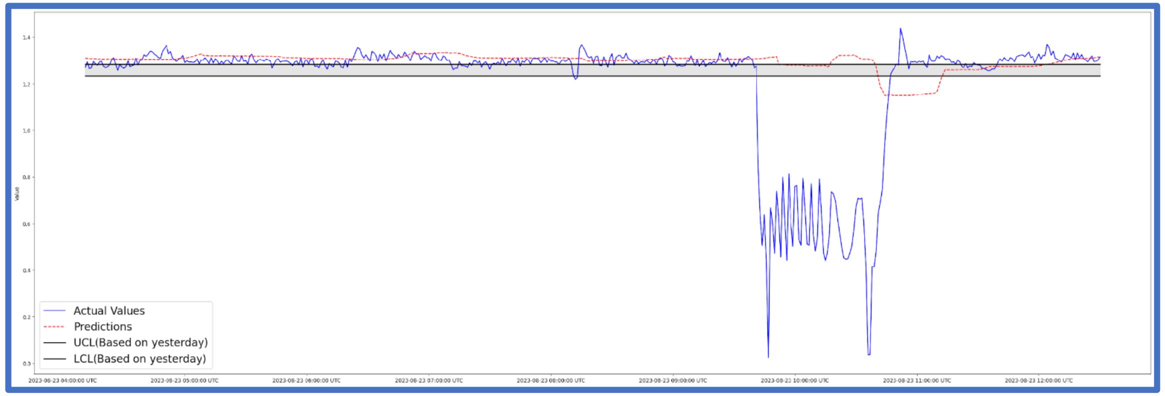
Industrial Artificial Intelligence Lab

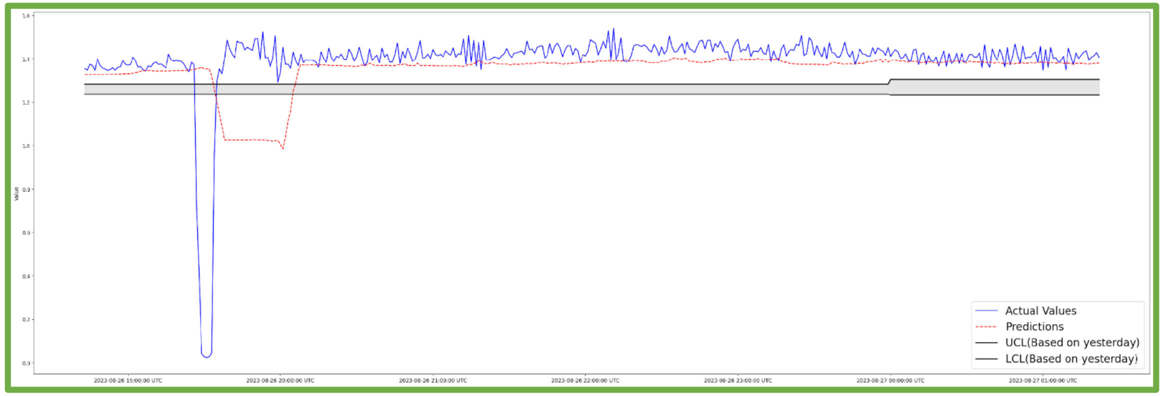
LSTM









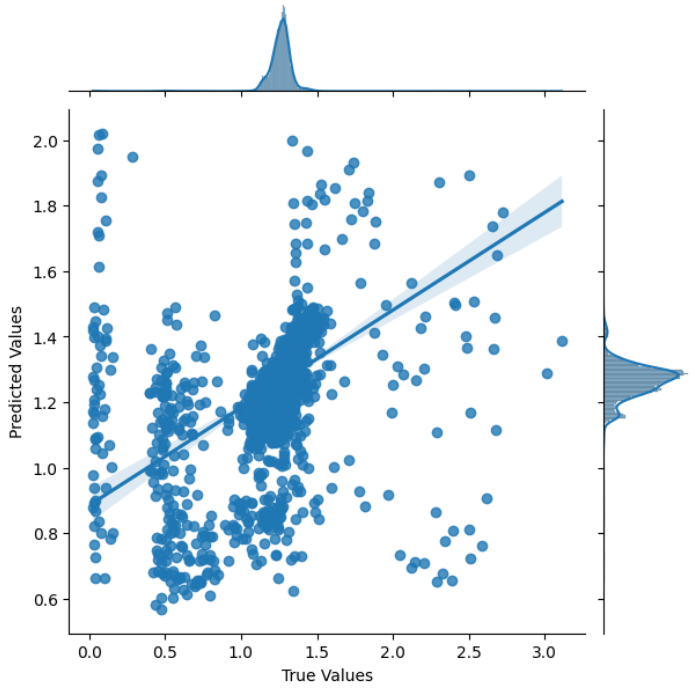


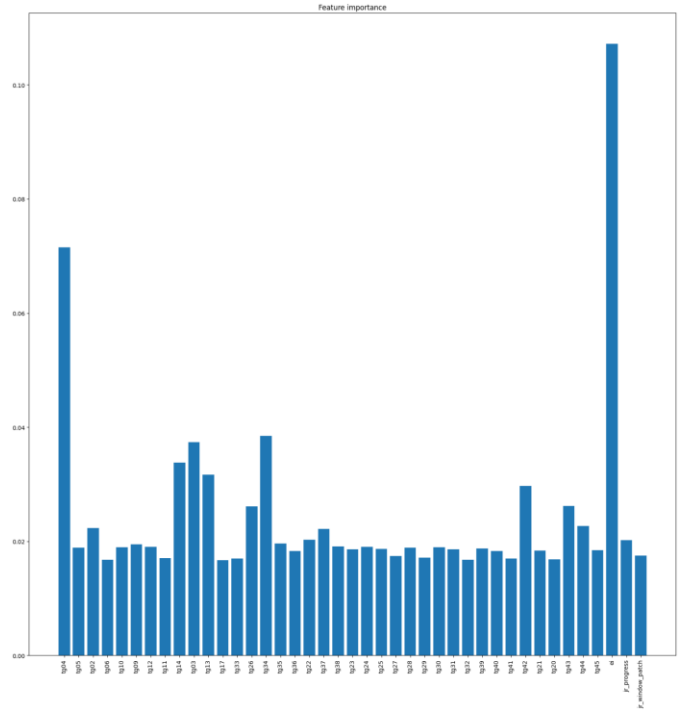
Industrial Artificial Intelligence Lab

IMV-LSTM

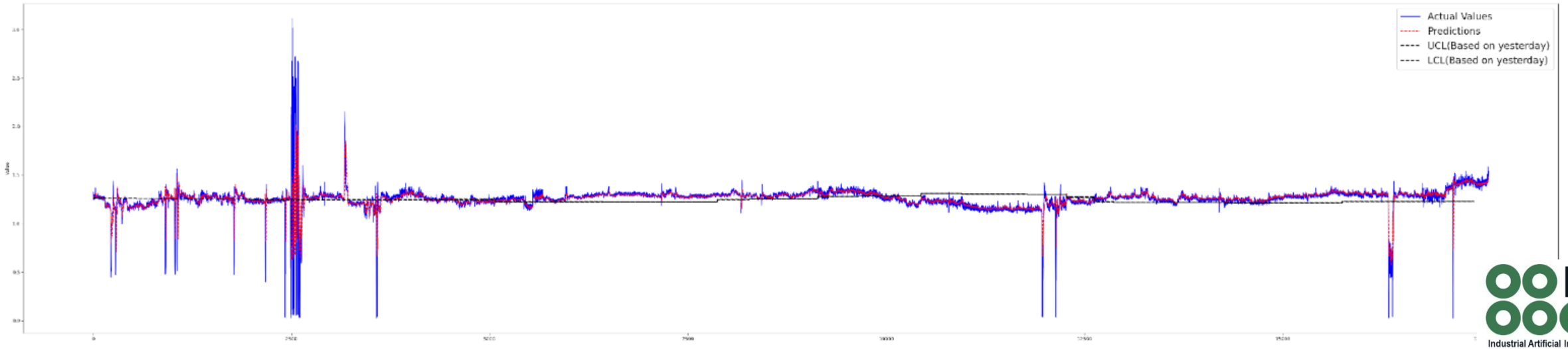
- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-3)
  - val loss가 가장 낮은 모델로 test
  - Lstm layer 노드 : 32

- Result
  - R^2 : 0.2052
  - MSE : 0.0134



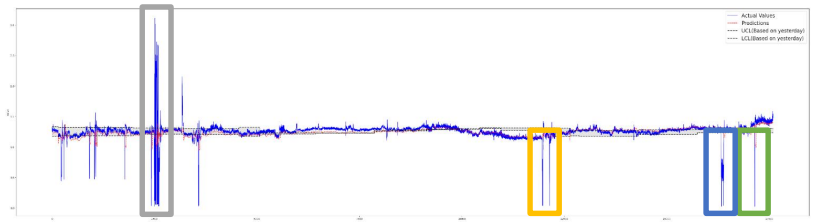


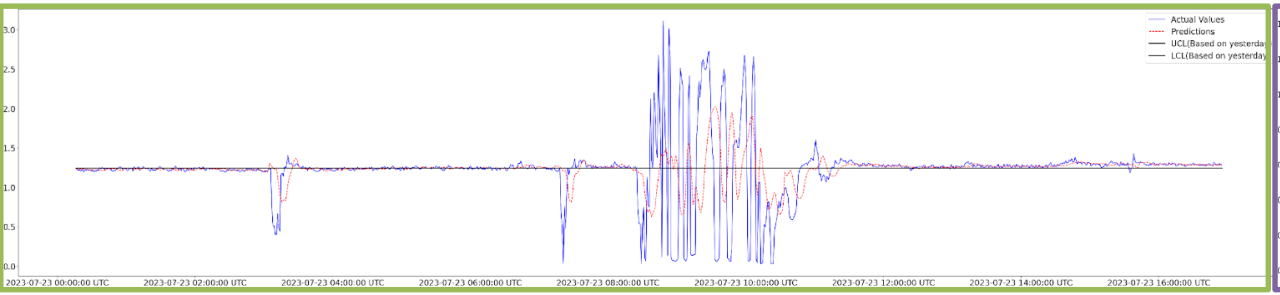
- 모델 해석(Attention Map)
  - ei, tg04

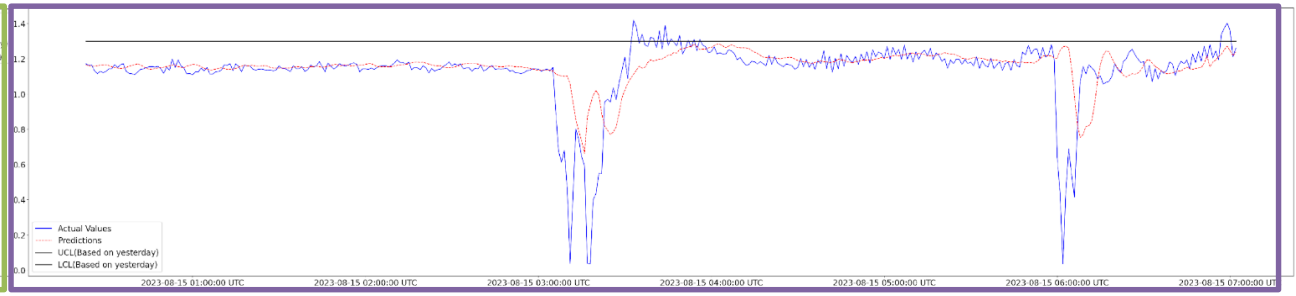


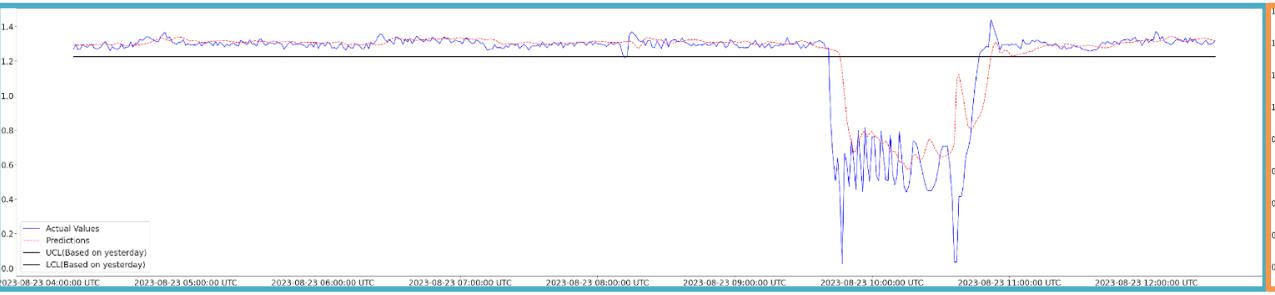
IAI
Industrial Artificial Intelligence Lab

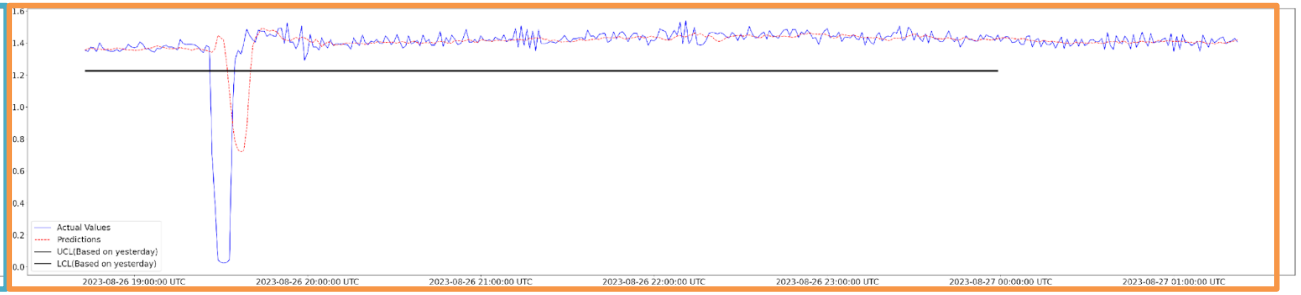
IMV-LSTM











Industrial Artificial Intelligence Lab

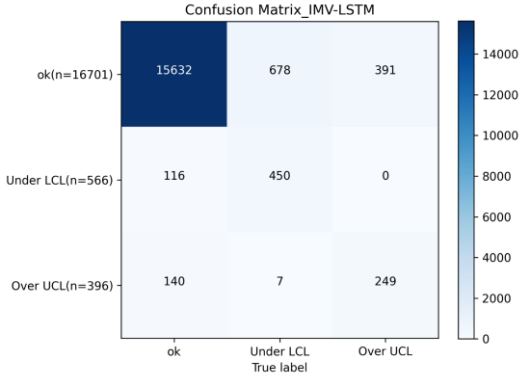
예측 설명 예시

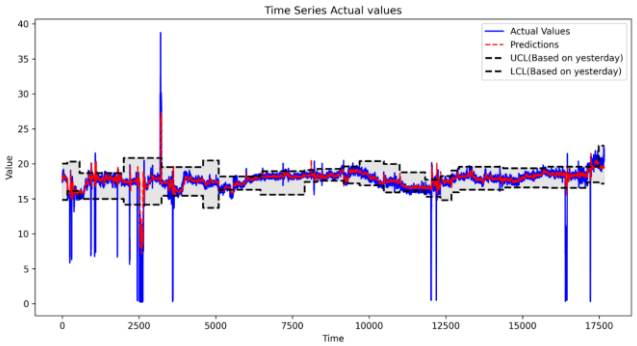
IMV-LSTM (Tg04에 대한 예측)

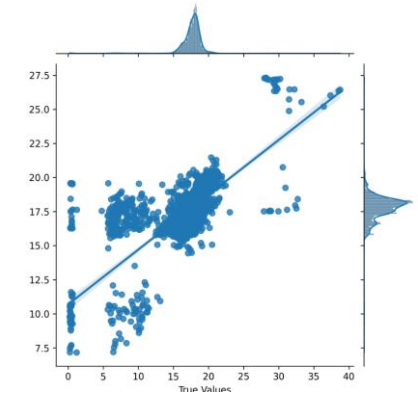
- Experiment setting
  - Epoch : 100
  - optimizer : Adam(lr=1e-3)
  - val loss가 가장 낮은 모델로 test
  - Lstm layer 노드 : 32

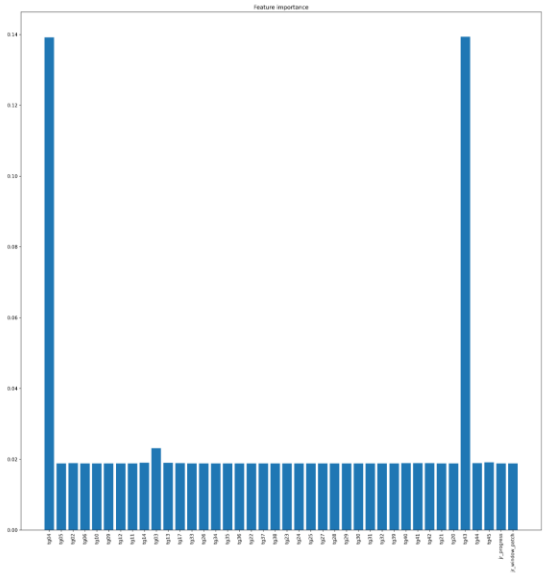
- Result
  - R^2 : 0.4881
  - MSE : 1.9021

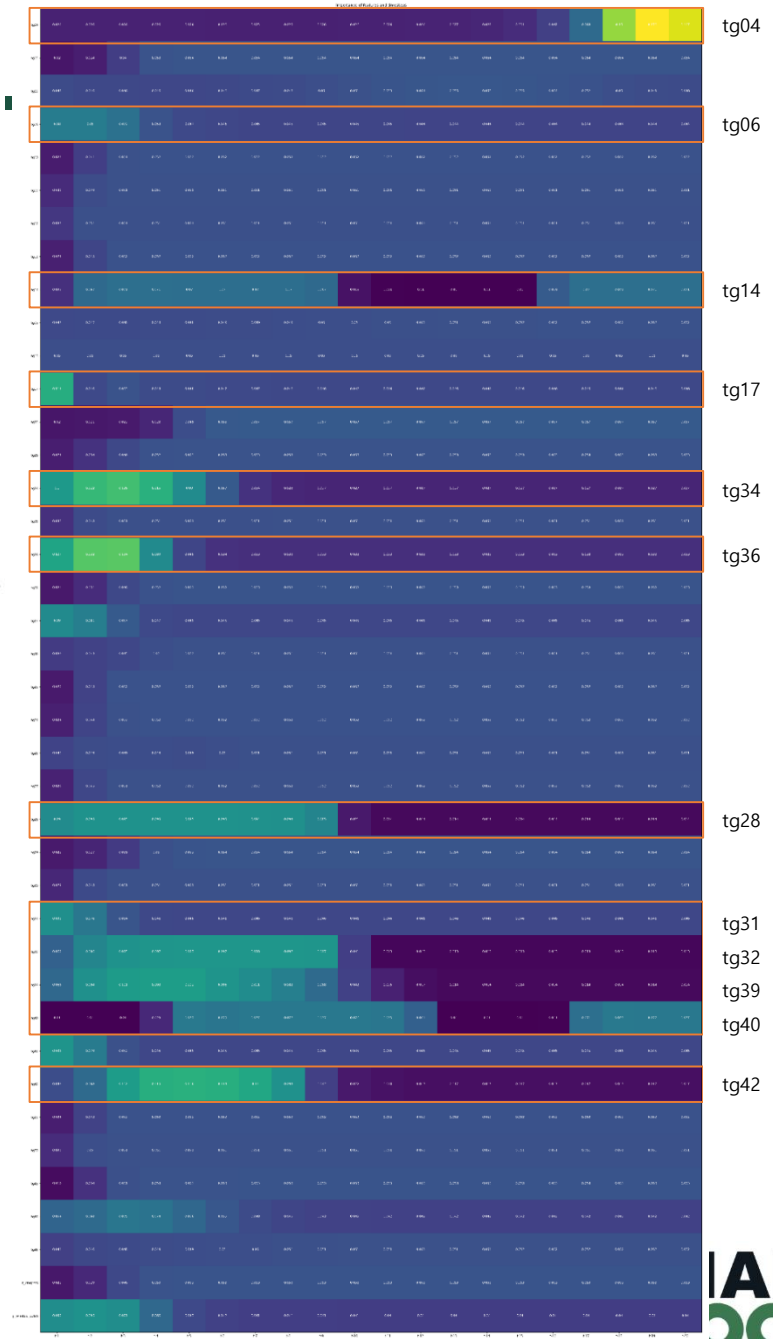
- **Global Attention Map**
  - tg04, tg43, tg03





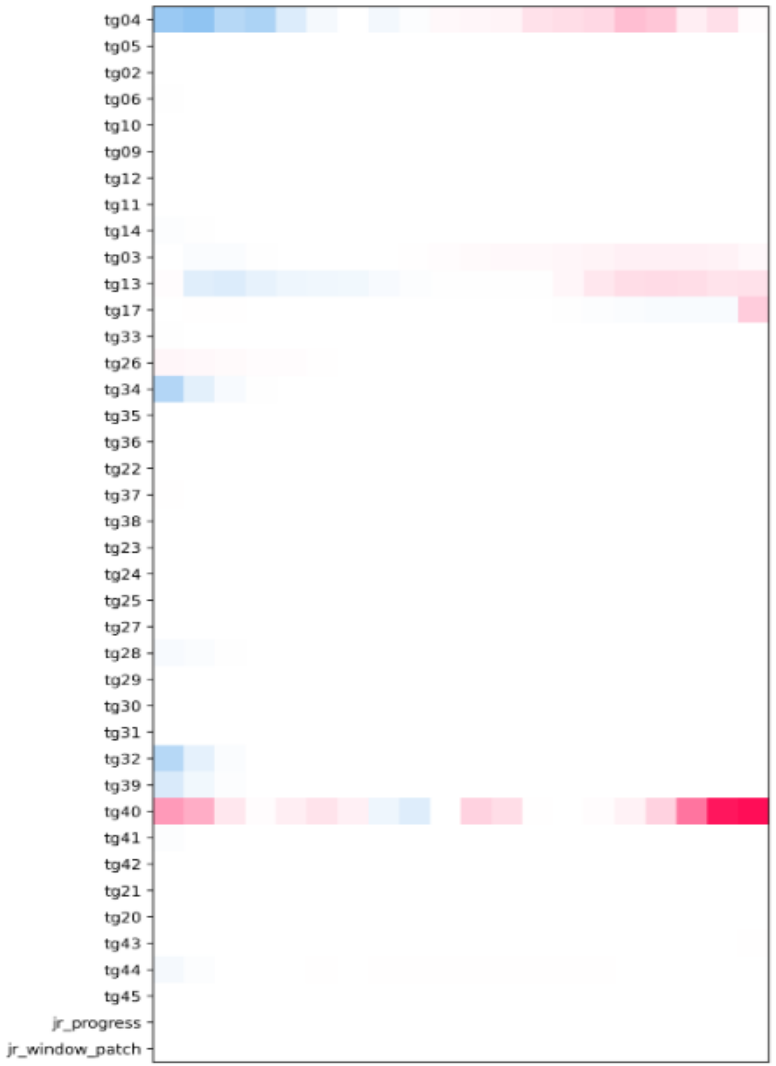


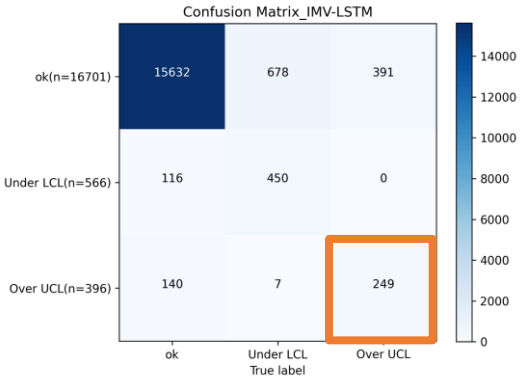


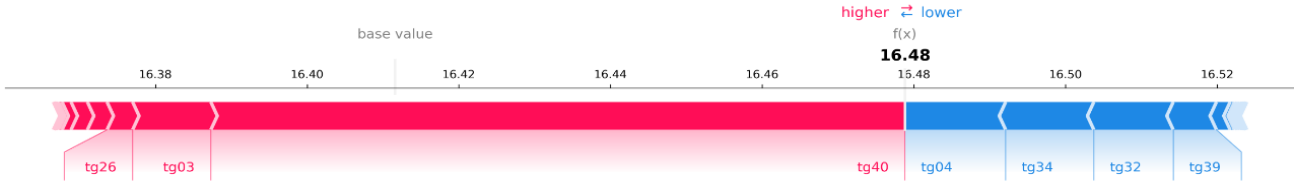


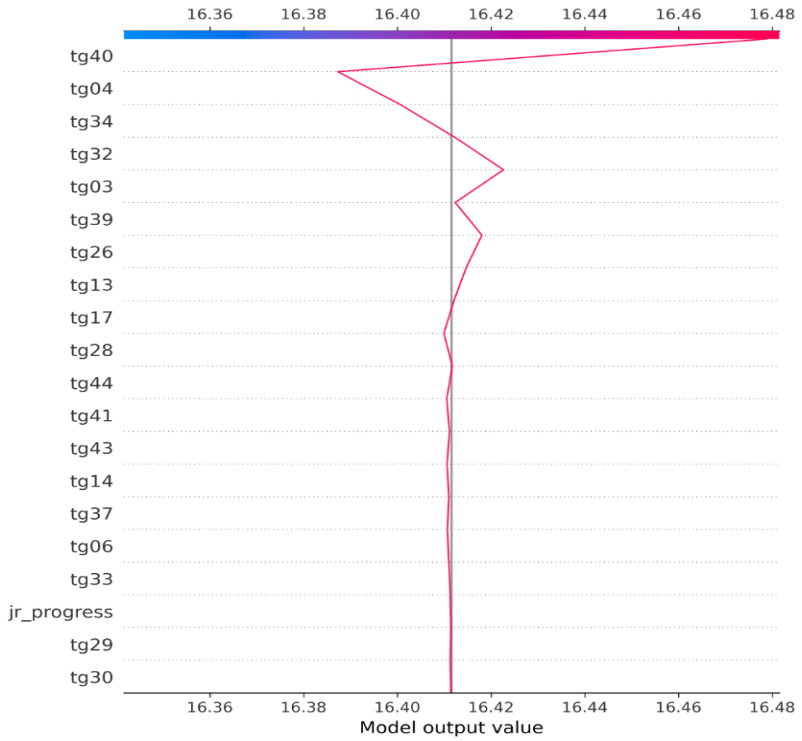
IMV-LSTM

- 개별 예측에 대한 해석 (local interpretation)
  SHAP - Correct_over (249)









Industrial Artificial Intelligence Lab

추후 계획

추후 계획

- Classification 문제로 변환하여 모델링 시도
- 코드 정리 / 모듈화
- 자료 작성

ETRI

SEOULTECH IAI Industrial Artificial Intelligence Lab

감사합니다