

24.01.25

이상치 및 주요 변수 정보

서울과학기술대학교 데이터사이언스학과

이성호 sean0310@seoultech.ac.kr

배소희 shbae2819@g.seoultech.ac.kr

심재웅 jaewoong@seoultech.ac.kr

- 이용 데이터

- 이용 데이터 정의

- input(X) : tg (38개 sensor 데이터), 공정진행도(jr_progress), 공정변화도(jr_window_patch), ei / 총 41개

- Window size : 30

- output(y) : 1분 후의 ei 계산값 ($tg04 / tg02 * tg03 * 0.0003$)

- Data split

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

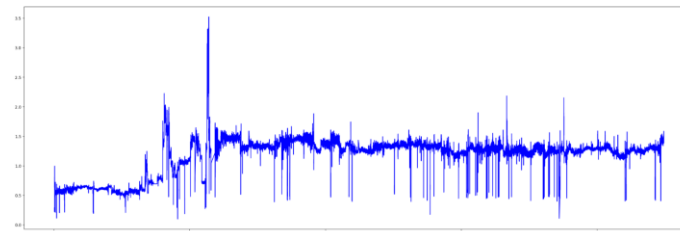
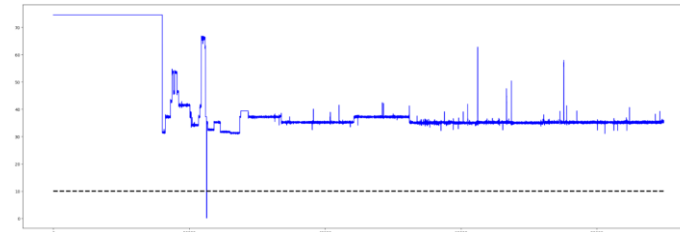
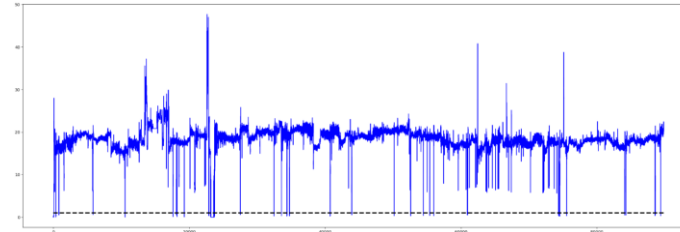
- Train : 2023-03-02 08:00:00 ~ 2023-06-16 18:37:00 (52830)

- Validation : ~ 2023-07-16 12:47:00 (17610)

- Test : ~ 2023-08-27 03:00:00 (17598)

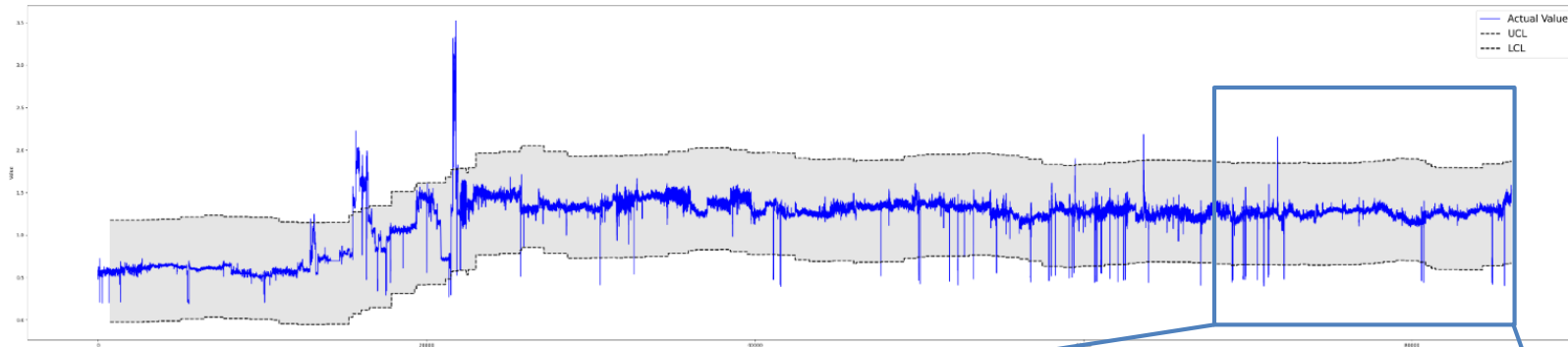
- tg04
Tg02 값 중 1 이하의 값 제외(735개의 index)
- tg03
Tg03 값 중 1000 이하의 값 제외(959개의 index)
- tg02
Tg02 값 중 10 이하의 값 제외(54개의 index)
- 결과
총 1,131개의 row가 삭제됨

```
count    88709.000000  
mean      1.174082  
std       0.305062  
min       0.096630  
25%      1.158493  
50%      1.274723  
75%      1.343347  
max       3.523615  
Name: ei, dtype: float64
```

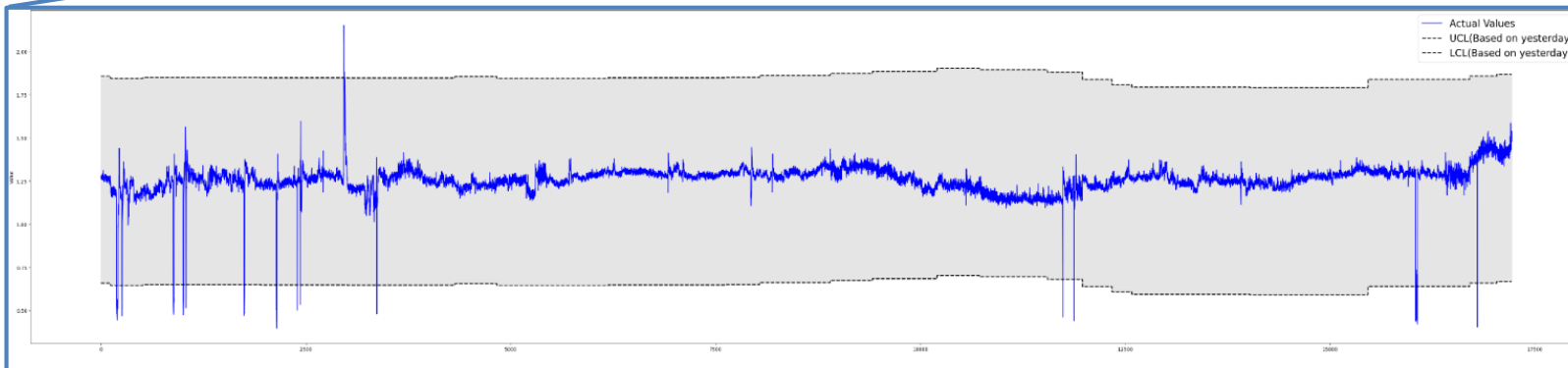


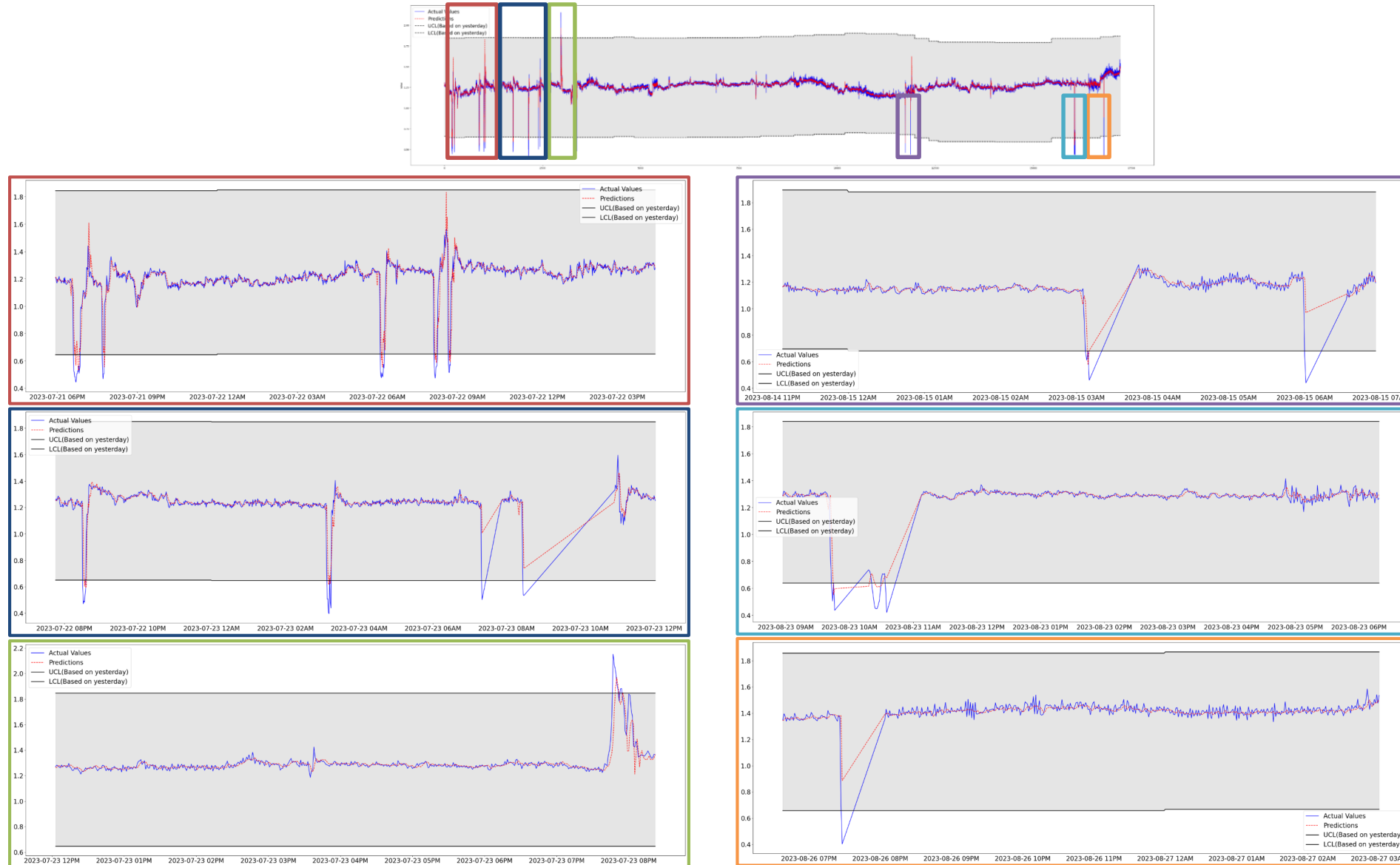
- 현재 날짜의 5일 이전의 $\text{mean} \pm 0.6$

- 전체 시계열

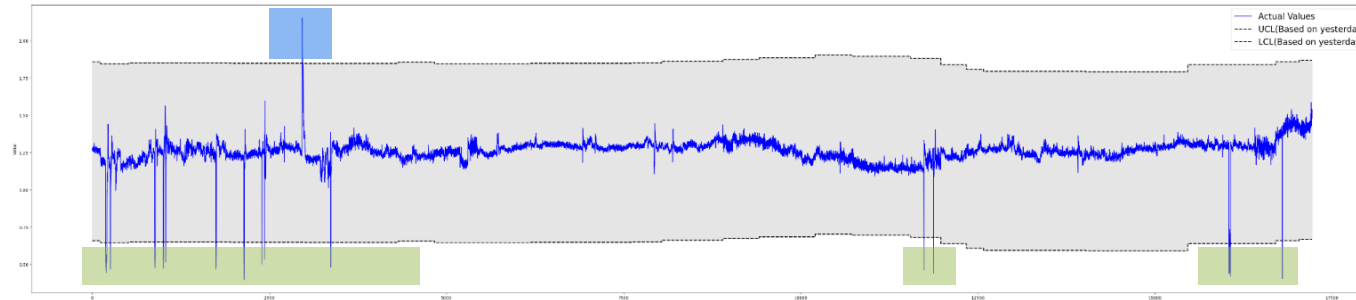


- Test 구간의 시계열





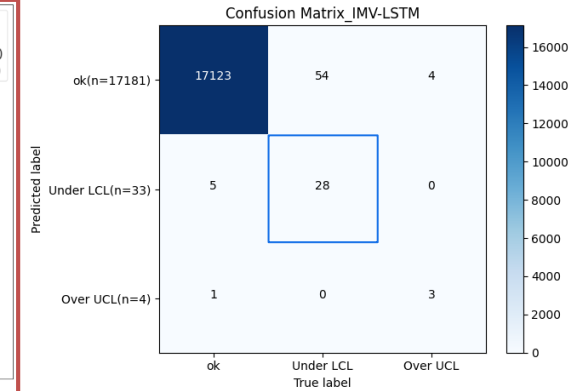
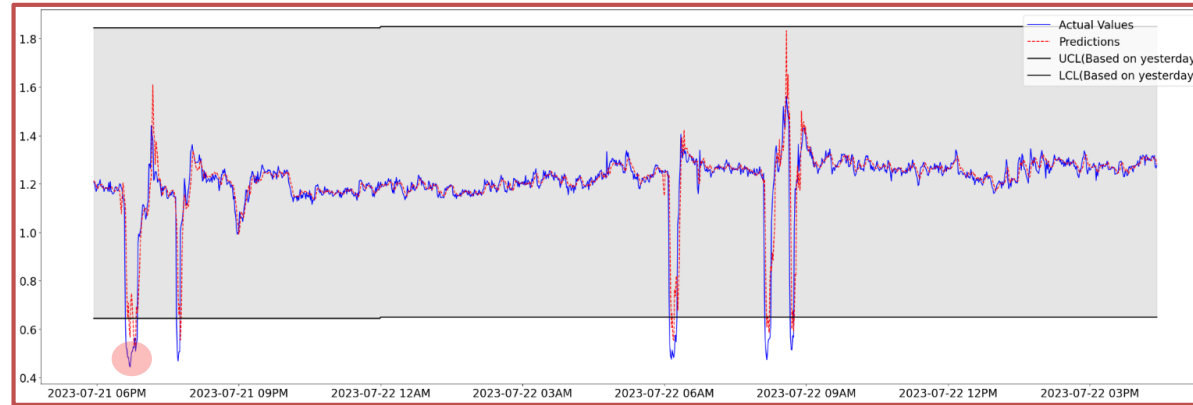
- 이상치에 해당하는 index (89개)
- 연속되는 값들 중 가장 낮거나 가장 높은 값 (19개)



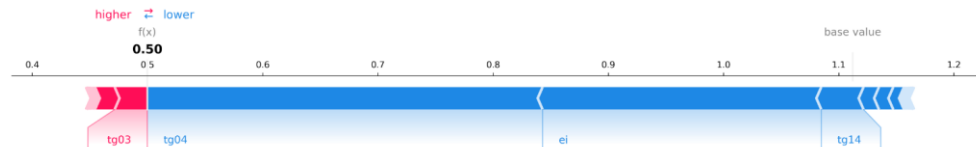
Index	date	ei
196	2023-07-21 18:42:00	0.4438
257	2023-07-21 19:43:00	0.4686
883	2023-07-22 6:09:00	0.4765
1004	2023-07-22 8:10:00	0.4737
1036	2023-07-22 8:42:00	0.5142
1745	2023-07-22 20:31:00	0.4711
2145	2023-07-23 3:11:00	0.3970
2148	2023-07-23 3:14:00	0.4393
2394	2023-07-23 7:20:00	0.5023
2431	2023-07-23 8:28:00	0.5335
2964	2023-07-23 19:47:00	2.1524
2970	2023-07-23 19:53:00	1.8793
3366	2023-07-25 7:25:00	0.4797
11742	2023-08-15 3:10:00	0.4606
11875	2023-08-15 6:01:00	0.4402
16049	2023-08-23 9:46:00	0.4365
16057	2023-08-23 10:25:00	0.4474
16067	2023-08-23 10:35:00	0.4201
16800	2023-08-26 19:28:00	0.4037

● Index : 196

- 2023-07-21 18:42:00
- ei : 0.4438
- Correct under
Predicted: 1, True: 1

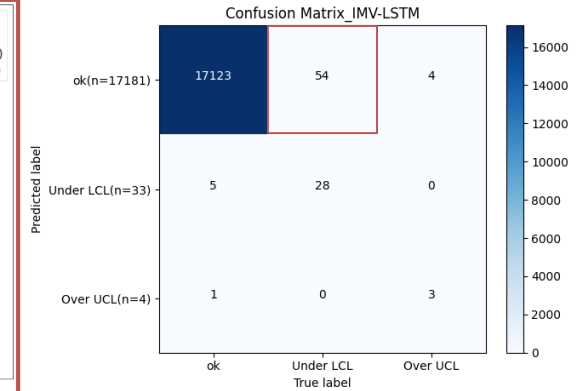
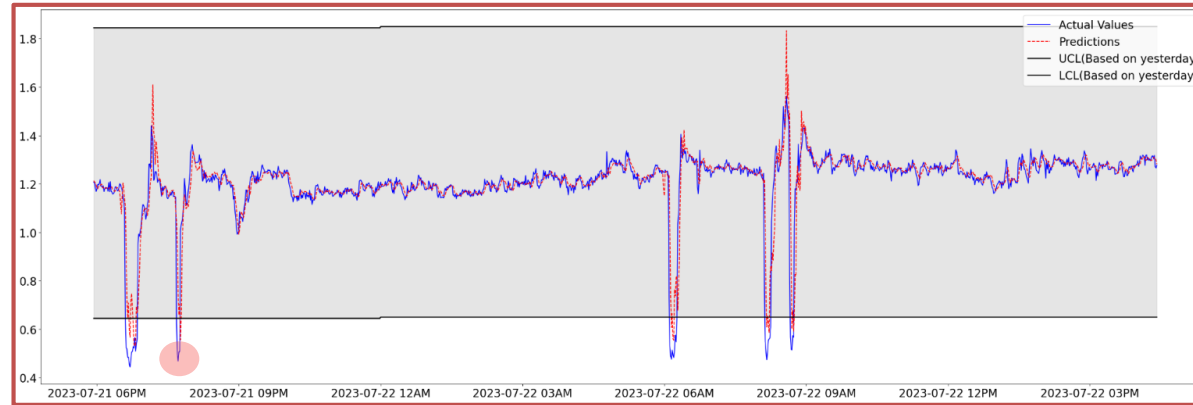


- 주요 영향 변수
 - tg04 : 스팀 순간값,
 - ei : 생산량,
 - tg14 : 설비B2 절대습도,
 - tg03 : 끝단 설비 속도,
 - tg02 : 종이별 측정 무게

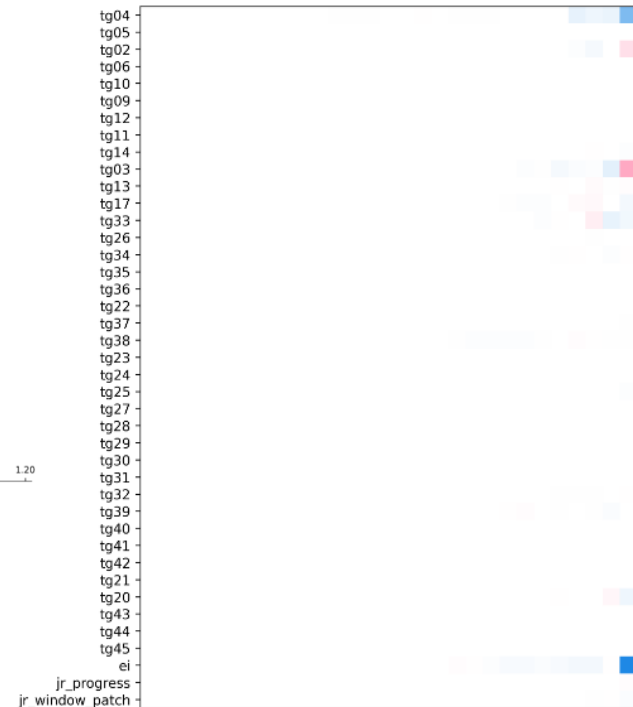
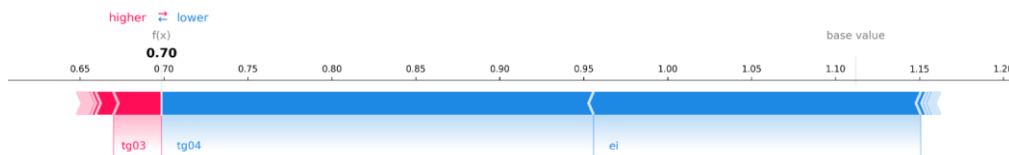


● Index : 257

- 2023-07-21 18:43:00
- ei : 0.4686
- Type1_error_under
Predicted: 0, True: 1

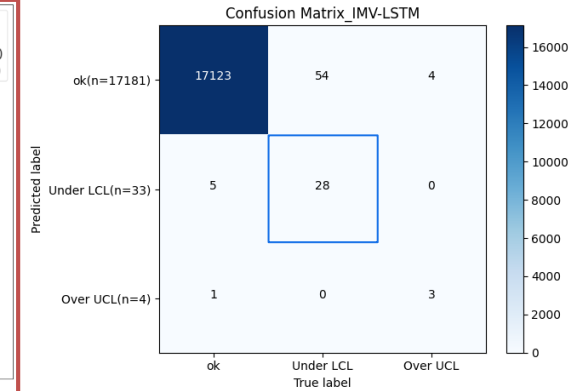
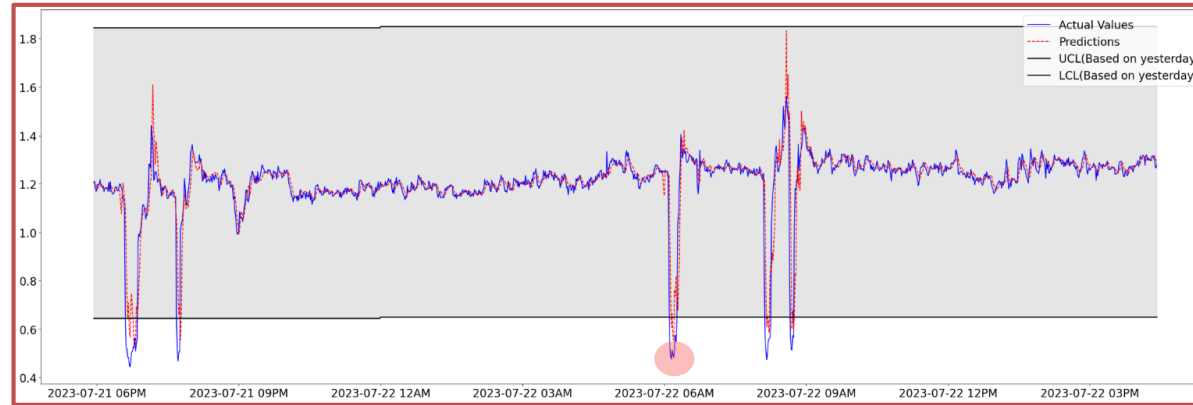


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg03 : 끝단 설비 속도,
tg02 : 종이별 측정 무게,
tg20 : 스팀 압력



● Index : 883

- 2023-07-22 06:09:00
- ei : 0.4765
- Correct under
Predicted: 1, True: 1

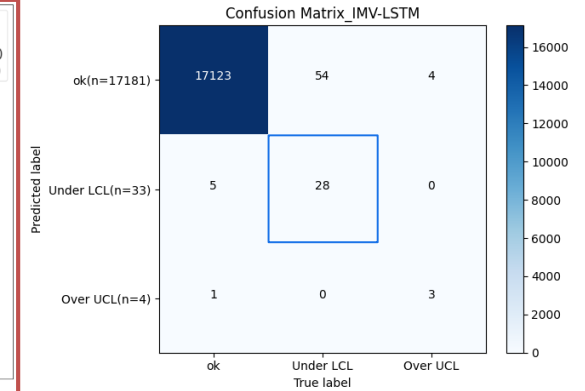
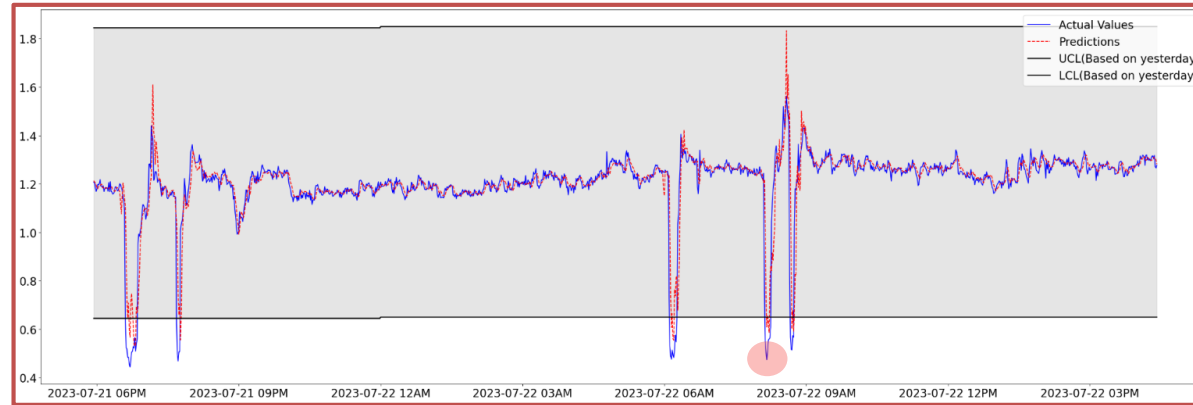


- 주요 영향 변수
 - tg04 : 스팀 순간값,
 - ei : 생산량,
 - tg02 : 종이별 측정 무게,
 - tg03 : 끝단 설비 속도,
 - tg17 : 스팀 누적값

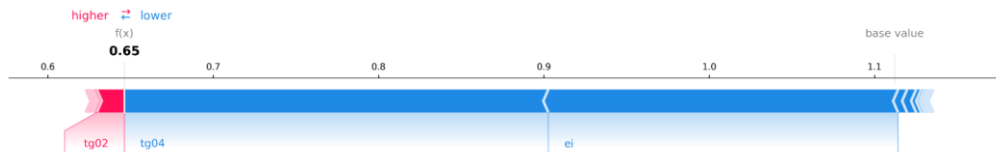


● Index : 1004

- 2023-07-22 08:10:00
- ei : 0.4737
- Correct under
Predicted: 1, True: 1

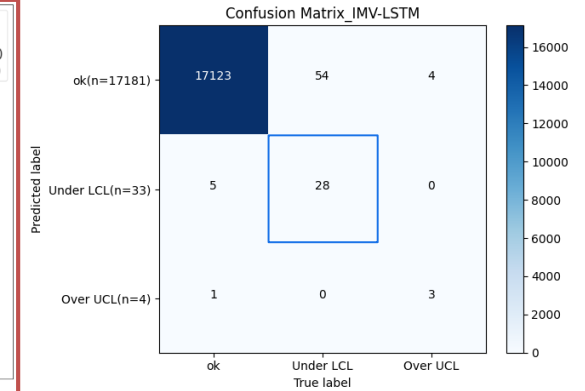
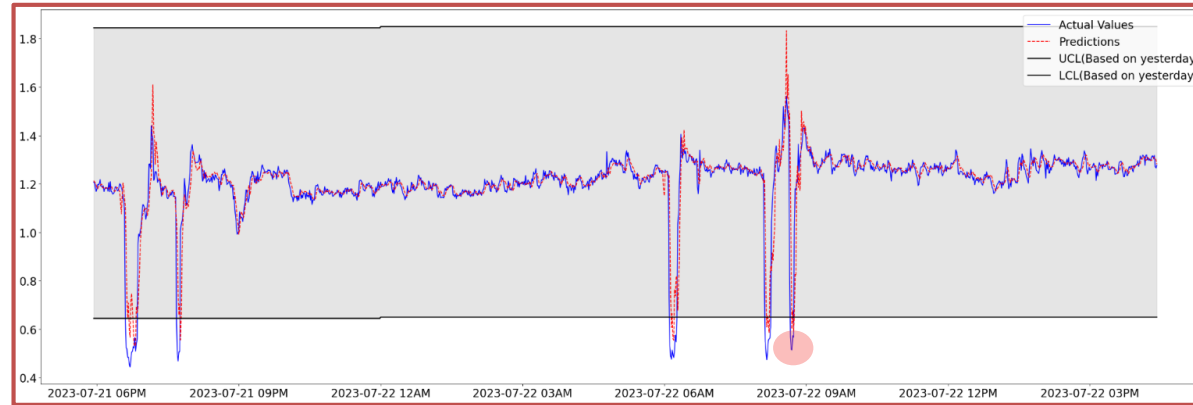


- 주요 영향 변수
 - tg04 : 스팀 순간값
 - ei : 생산량,
 - tg02 : 종이별 측정 무게,
 - tg03 : 끝단 설비 속도,
 - tg38 : 설비 AE 속도

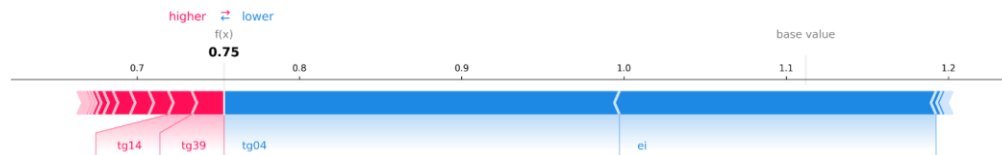


● Index : 1036

- 2023-07-22 08:42:00
- ei : 0.5142
- Correct under
Predicted: 1, True: 1

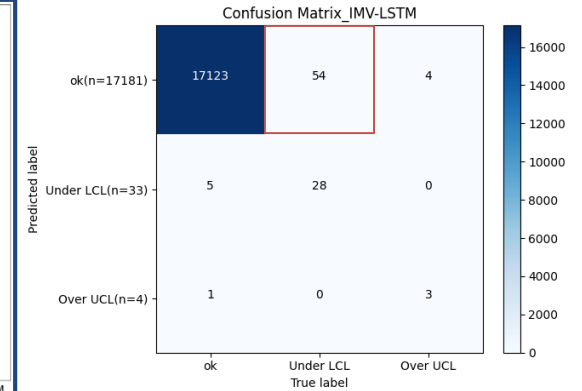
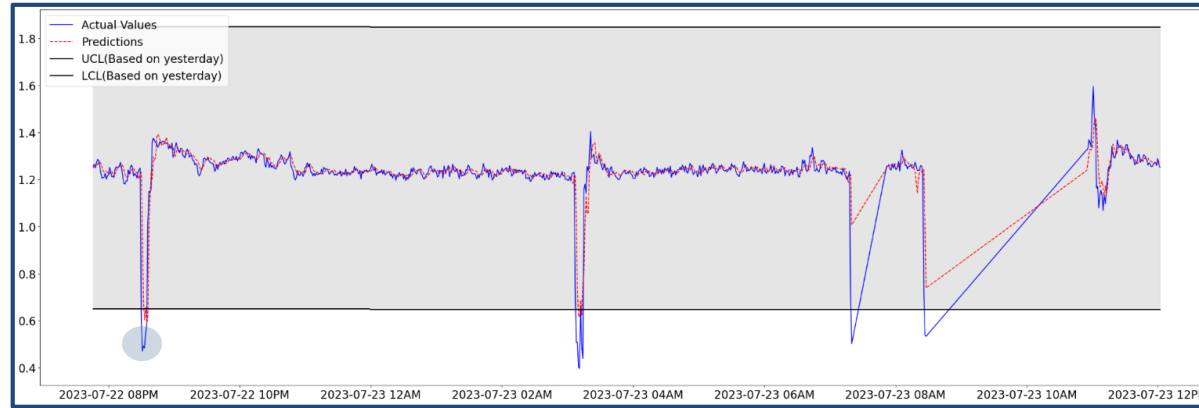


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg39 : 설비 AS 절대습도,
tg14 : 설비B2 절대습도,
tg02 : 종이별 측정 무게

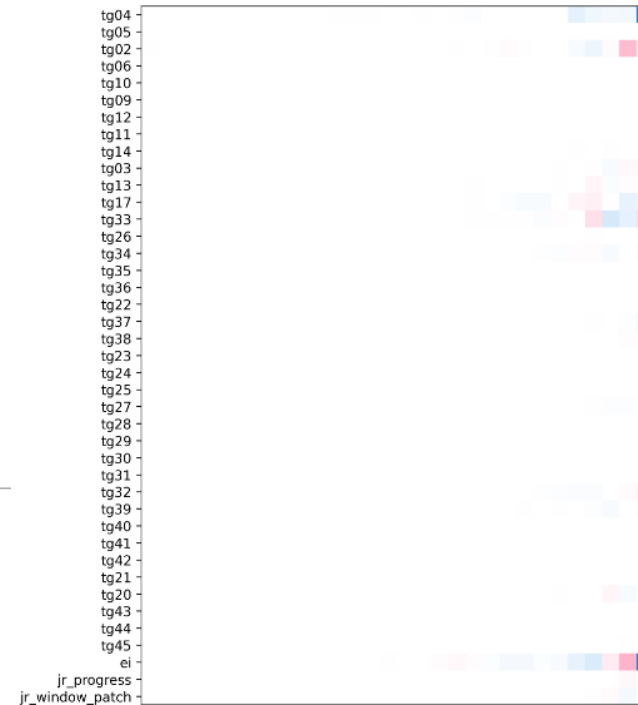
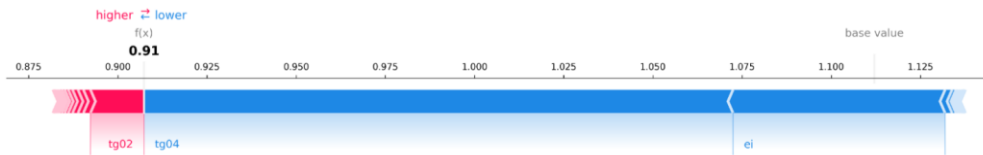


● Index : 1745

- 2023-07-22 20:31:00
- ei : 0.4711
- Type1_error_under
Predicted: 0, True: 1

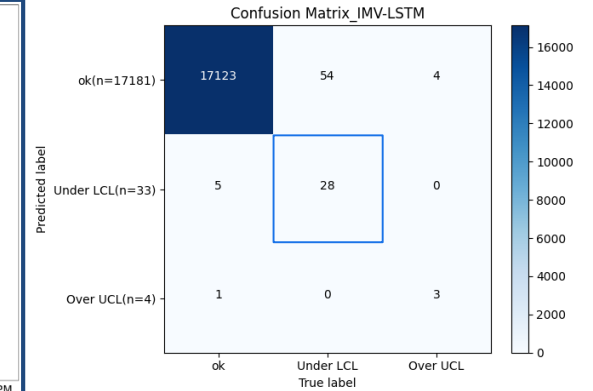
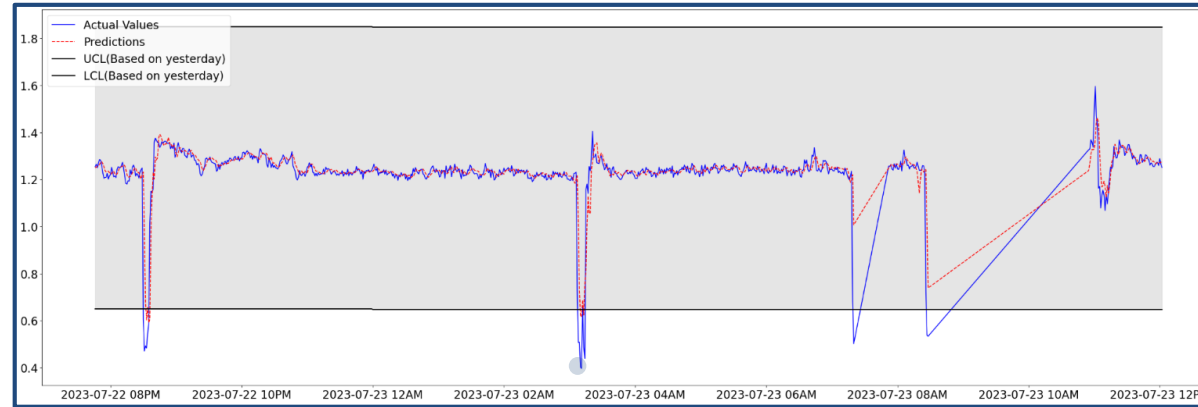


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg17 : 스팀 누적값,
tg37 : 설비 PE2 속도

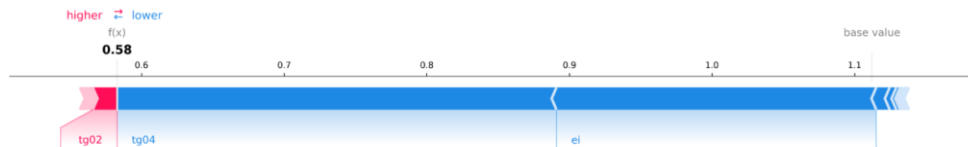


● Index : 2145

- 2023-07-23 03:11:00
- ei : 0.3970
- Correct_under
Predicted: 1, True: 1

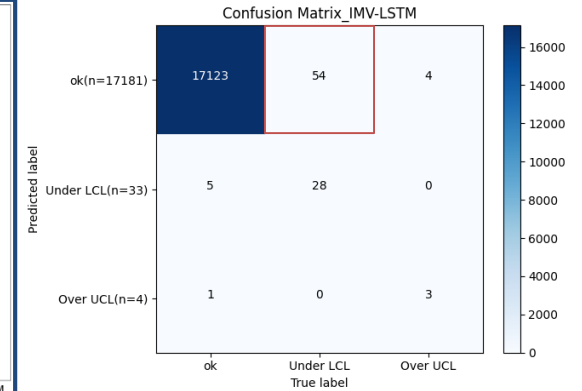
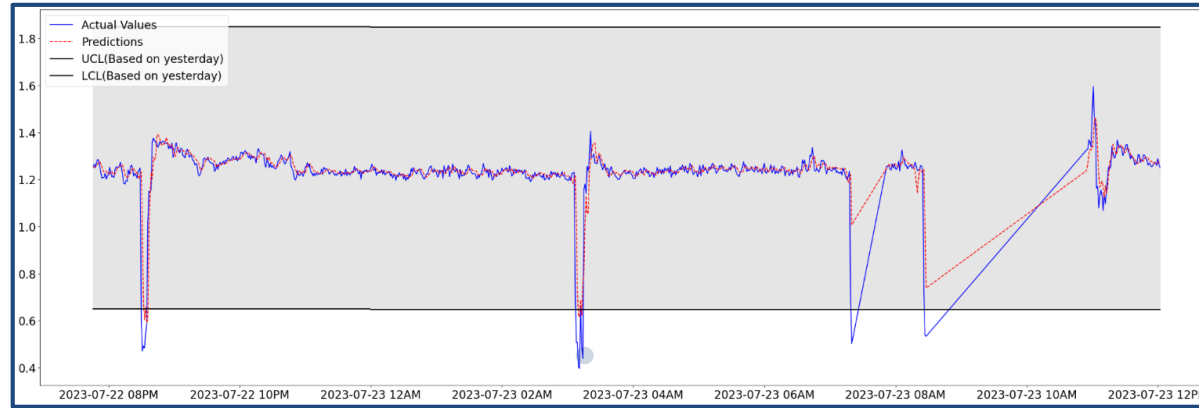


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg38 : 설비 AE 속도,
tg20 : 스팀 압력

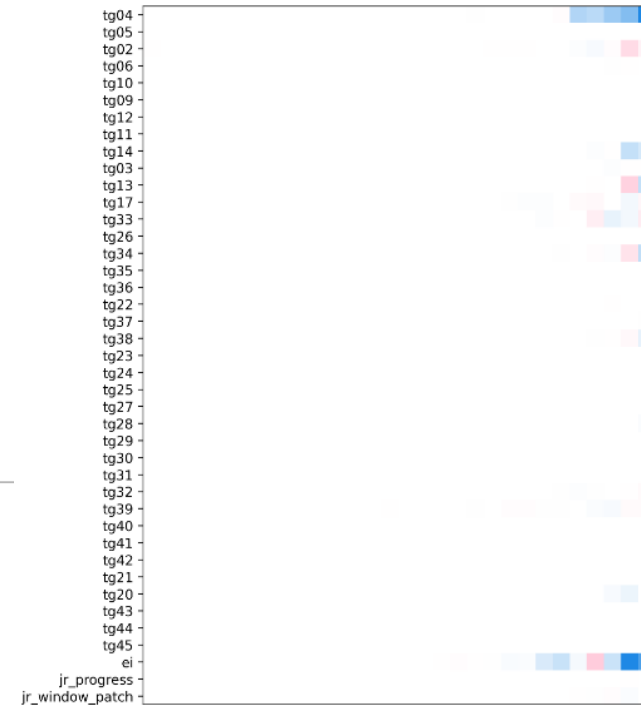
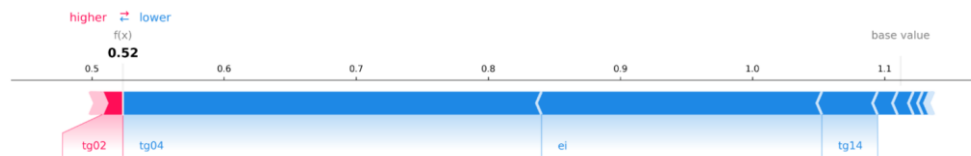


● Index : 2148

- 2023-07-23 03:14:00
- ei : 0.5142
- Type1_error_under
Predicted: 0, True: 1

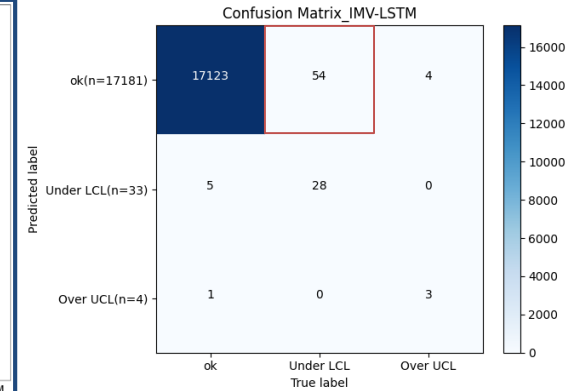
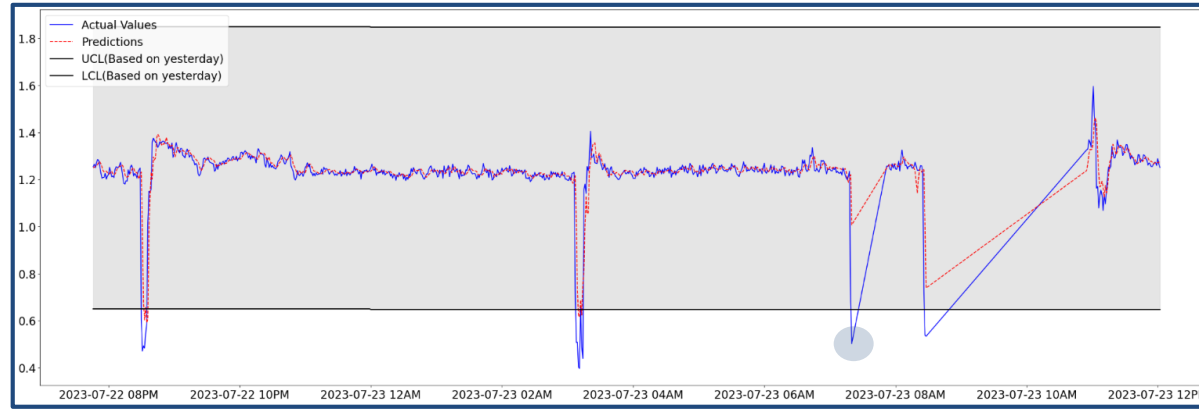


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg14 : 설비B2 절대습도,
tg02 : 종이별 측정 무게,
tg34 : 설비AE 절대습도

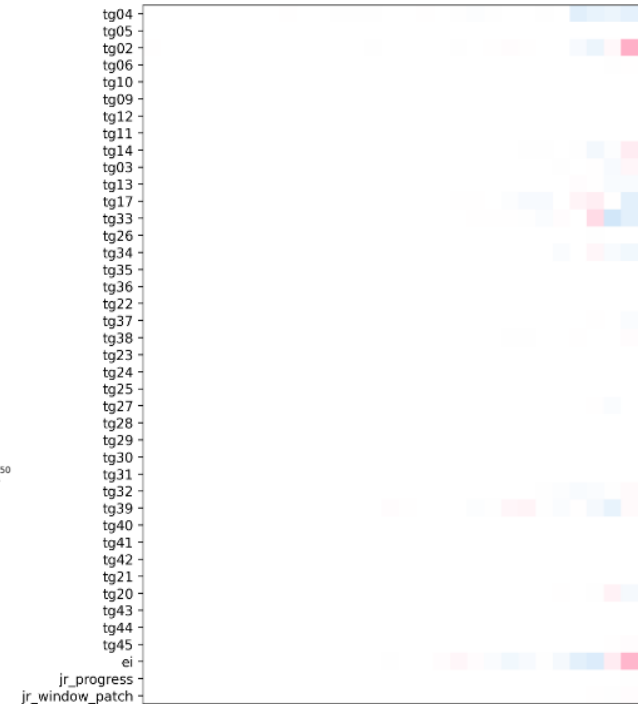
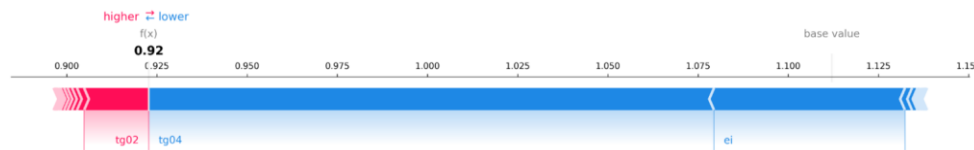


● Index : 2394

- 2023-07-23 07:20:00
- ei : 0.5023
- Type1_error_under
Predicted: 0, True: 1

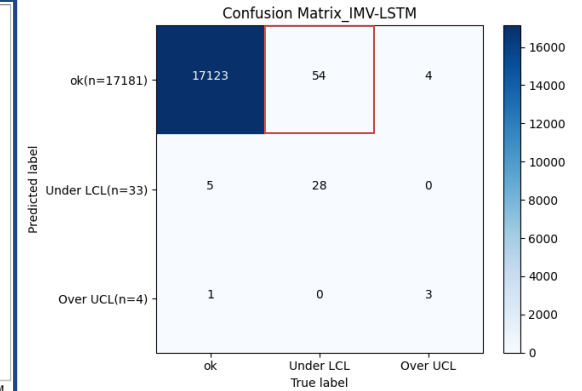
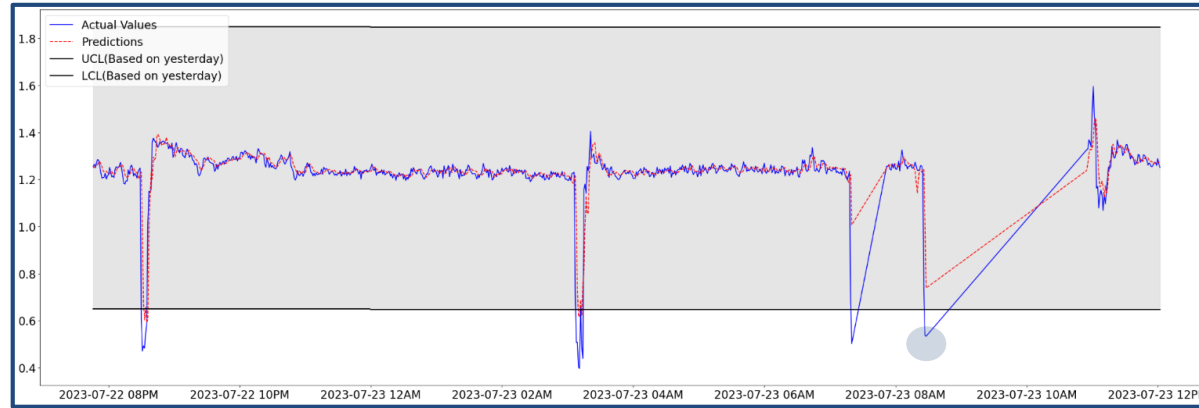


- 주요 영향 변수
tg04 : 스팀 순간값,
ei : 생산량,
tg02 : 종이별 측정 무게,
tg17 : 스팀 누적값,
tg37 : 설비 PE2 속도

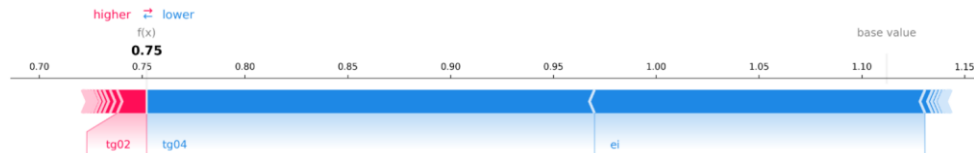


● Index : 2431

- 2023-07-23 08:28:00
- ei : 0.5335
- Type1_error_under
Predicted: 0, True: 1

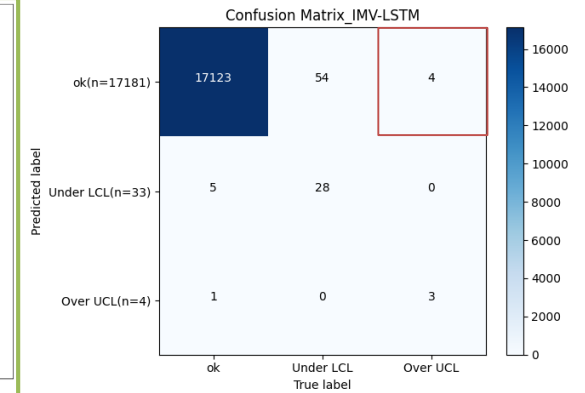
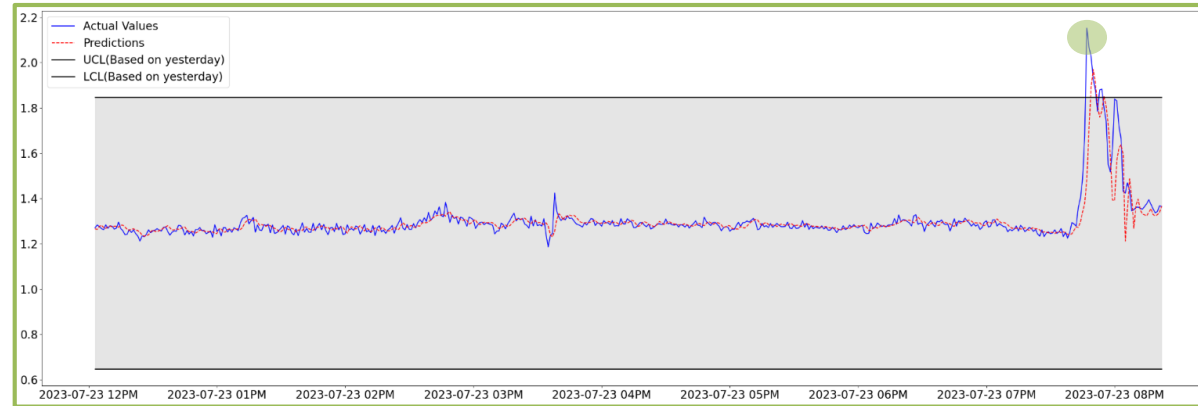


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg20 : 스팀 압력,
tg03 : 끝단 설비 속도



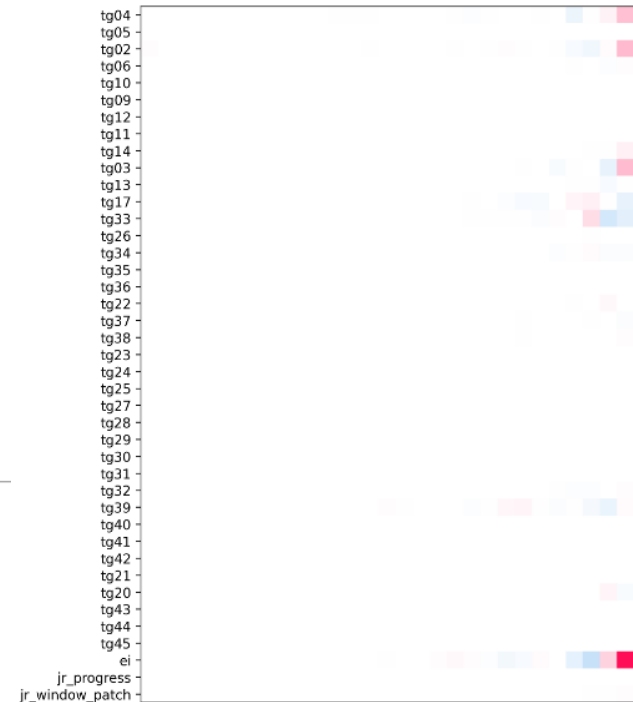
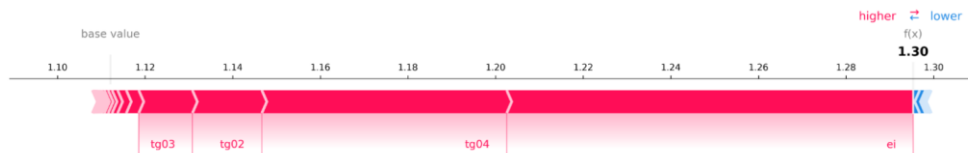
● Index : 2964

- 2023-07-23 19:47:00
- ei : 2.1524
- Type1_error_over
Predicted: 0, True: 2



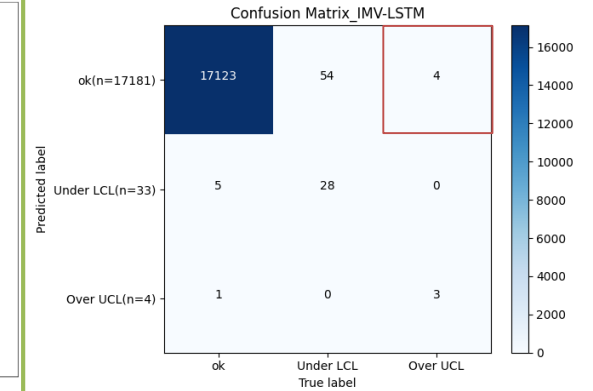
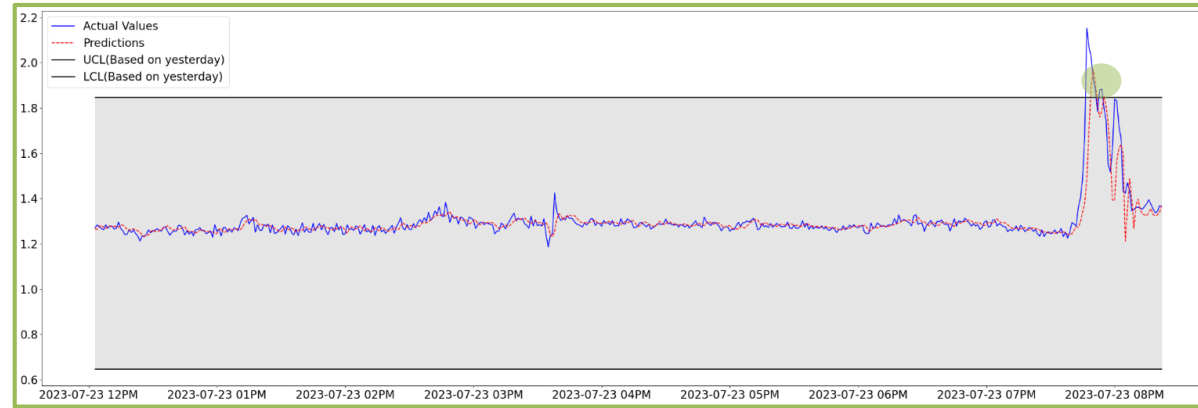
○ 주요 영향 변수

ei : 생산량,
tg04 : 스팀 순간값,
tg02 : 종이별 측정 무게,
tg03 : 끝단 설비 속도,
tg14 : 설비B2 절대속도

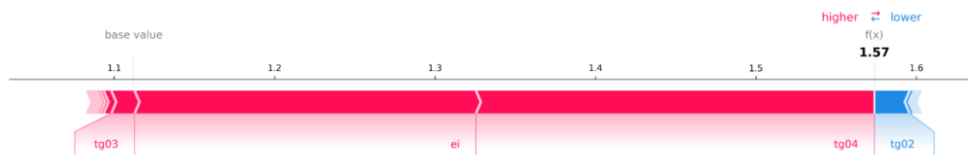


● Index : 2970

- 2023-07-23 19:53:00
- ei : 1.8793
- Type1_error_over
Predicted: 0, True: 2

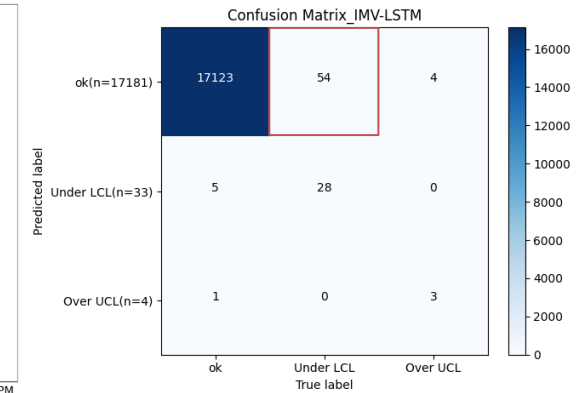
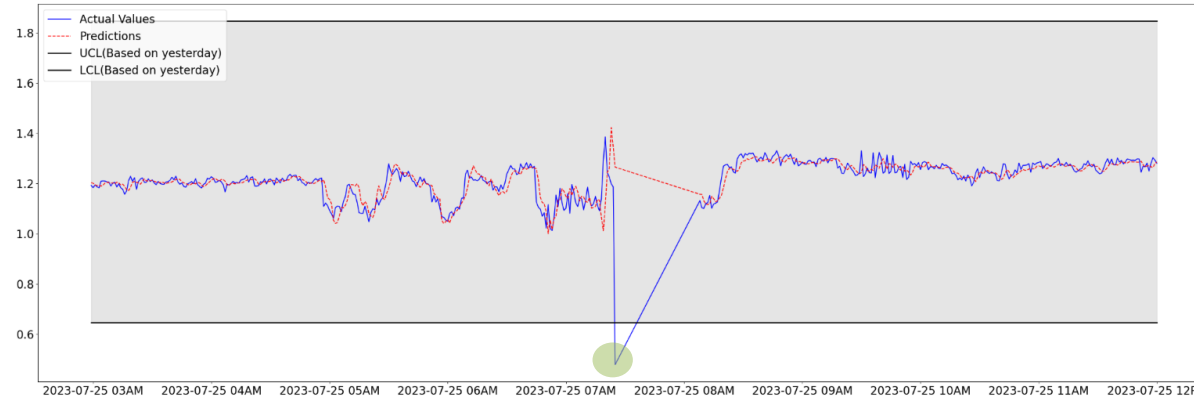


- 주요 영향 변수
 - tg04 : 스팀 순간값
 - ei : 생산량,
 - tg02 : 종이별 측정 무게,
 - tg03 : 끝단 설비 속도,
 - tg14 : 설비B2 절대습도

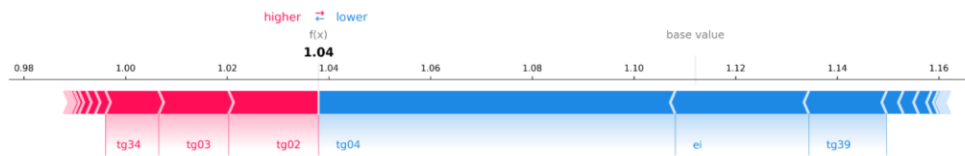


● Index : 3366

- 2023-07-25 07:25:00
- ei : 0.4797
- Type1_error_under
Predicted: 0, True: 1

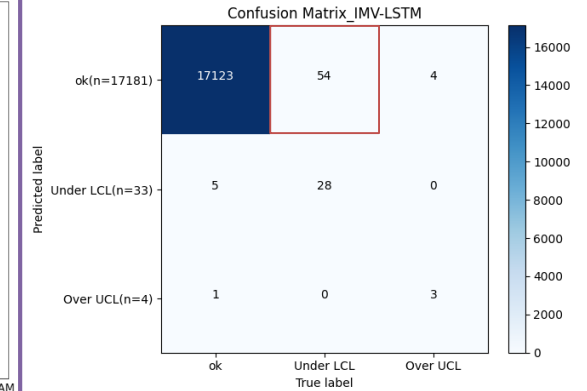
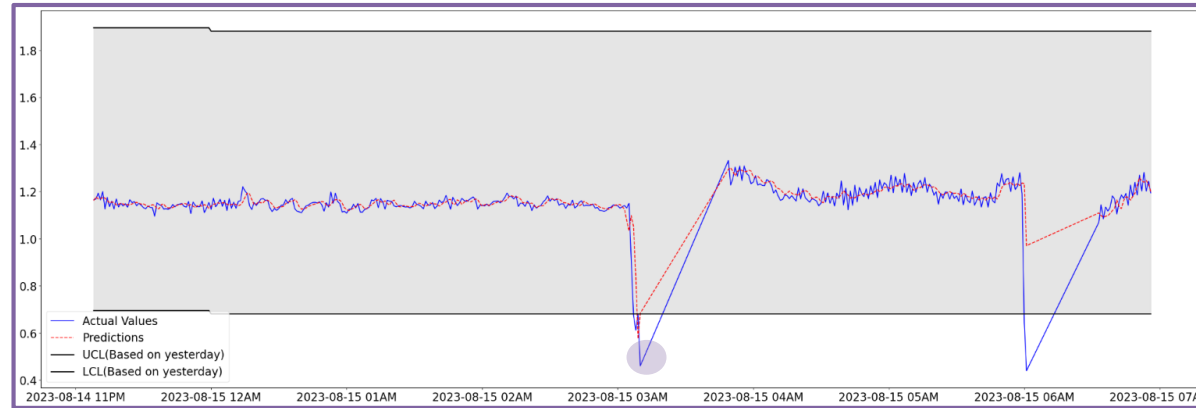


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg39 : 설비 AS 절대습도,
tg03 : 끝단 설비 속도

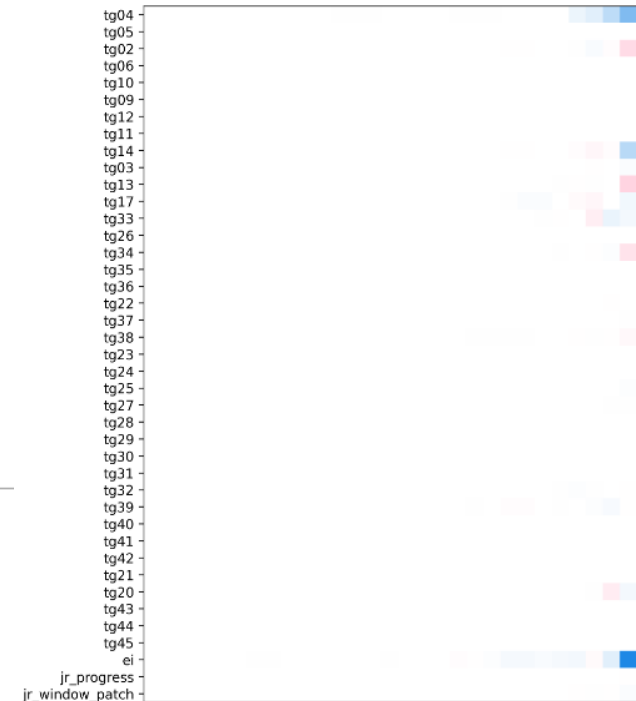


● Index : 11742

- 2023-08-15 03:10:00
- ei : 0.4606
- Type1_error_under
Predicted: 0, True: 1

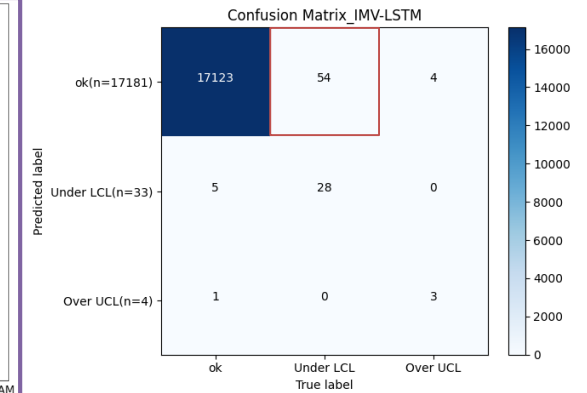
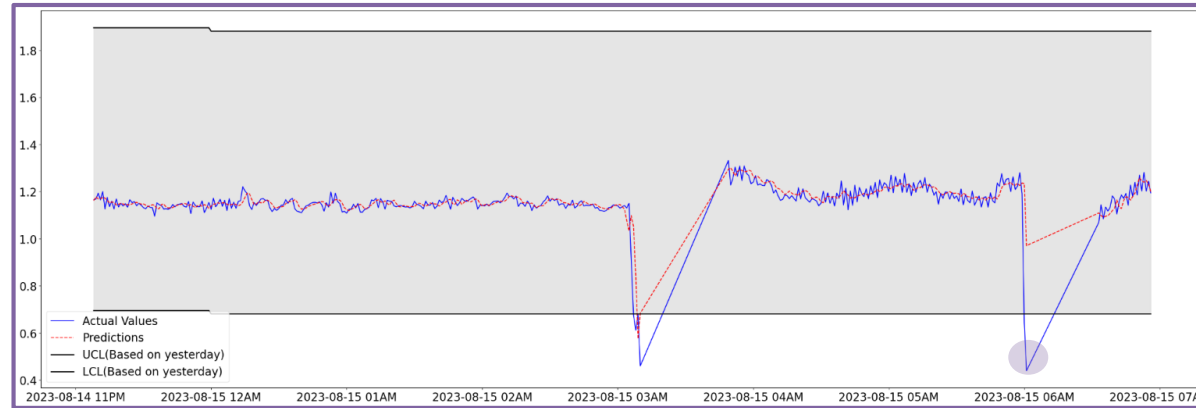


- 주요 영향 변수
 - tg04 : 스팀 순간값,
 - ei : 생산량,
 - tg14 : 설비B2 절대습도,
 - tg02 : 종이별 측정 무게,
 - tg34 : 설비AE 절대습도

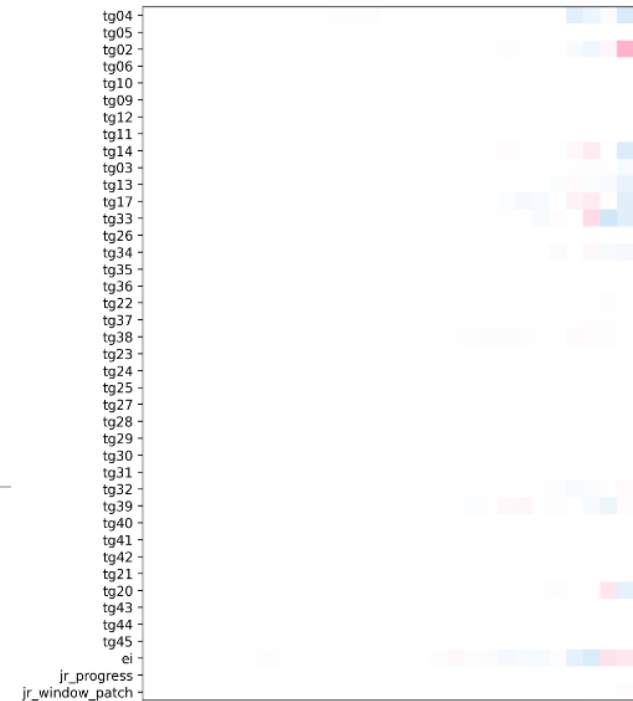
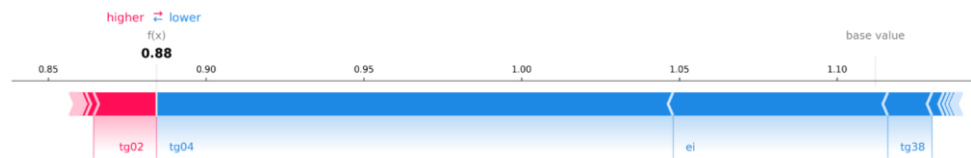


● Index : 11875

- 2023-08-15 16:01:00
- ei : 0.4402
- Type1_error_under
Predicted: 0, True: 1

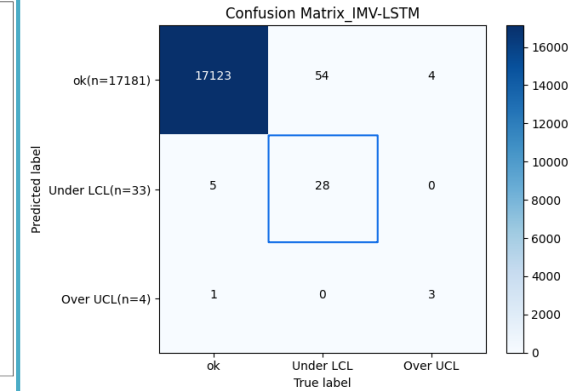
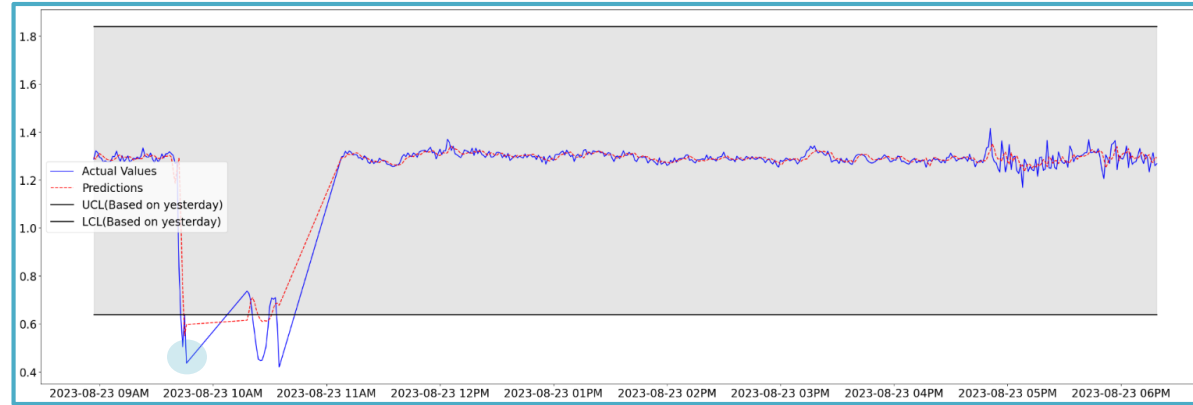


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg38 : 설비 AE 속도,
tg03 : 끝단 설비 속도

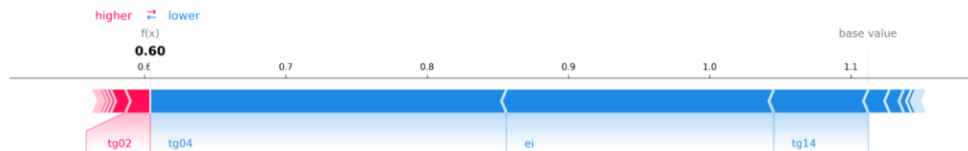


● Index : 16049

- 2023-08-23 09:46:00
- ei : 0.4365
- Correct under
Predicted: 1, True: 1

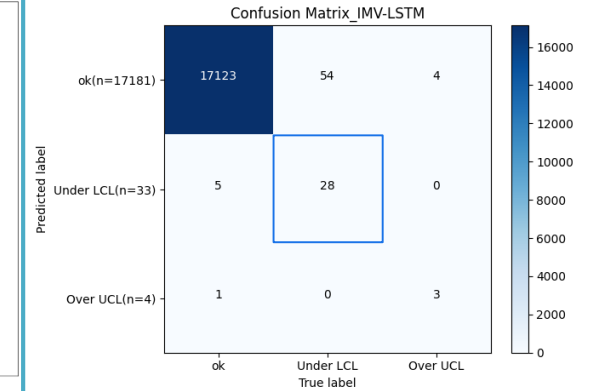
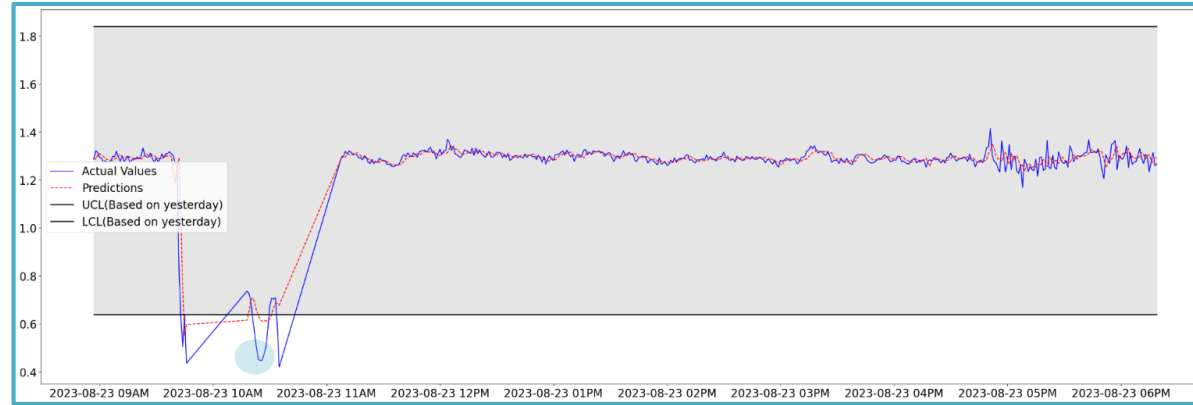


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg14 : 설비B2 절대습도,
tg02 : 종이별 측정 무게,
tg34 : 설비AE 절대습도

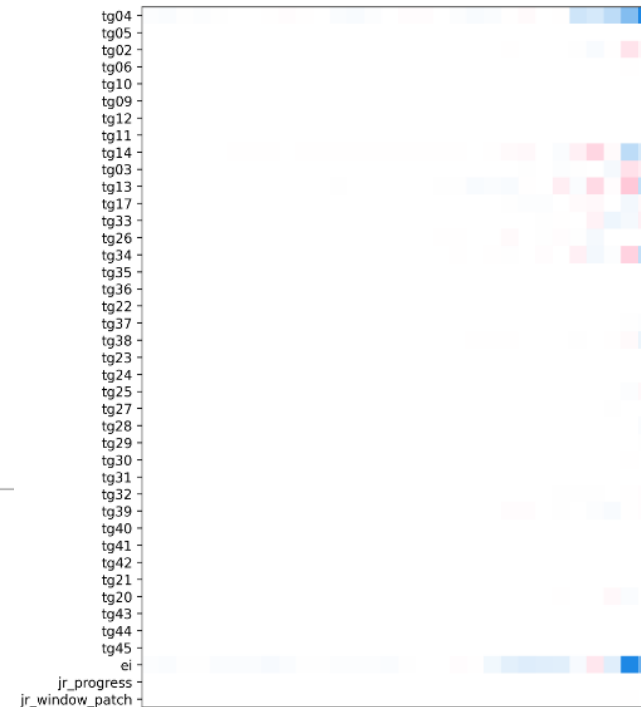
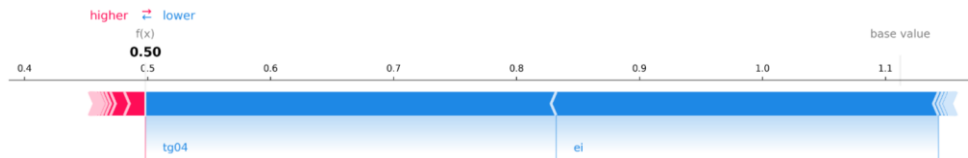


● Index : 16057

- 2023-08-23 10:25:00
- ei : 0.4474
- Correct under
Predicted: 1, True: 1

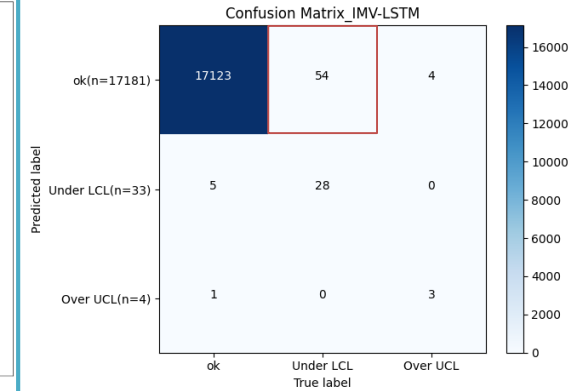
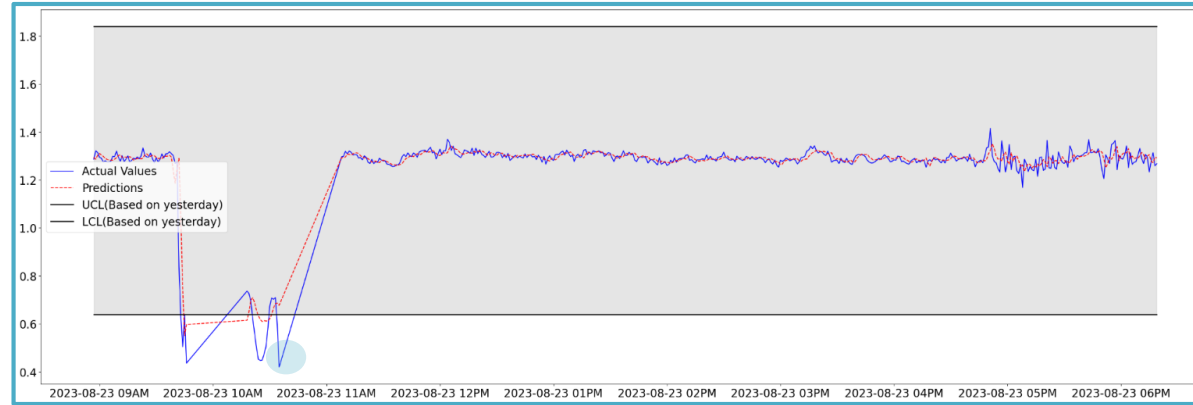


- 주요 영향 변수
tg04 : 스팀 순간값
ei : 생산량,
tg02 : 종이별 측정 무게,
tg03 : 끝단 설비 속도,
tg25 : 설비 PS1 속도

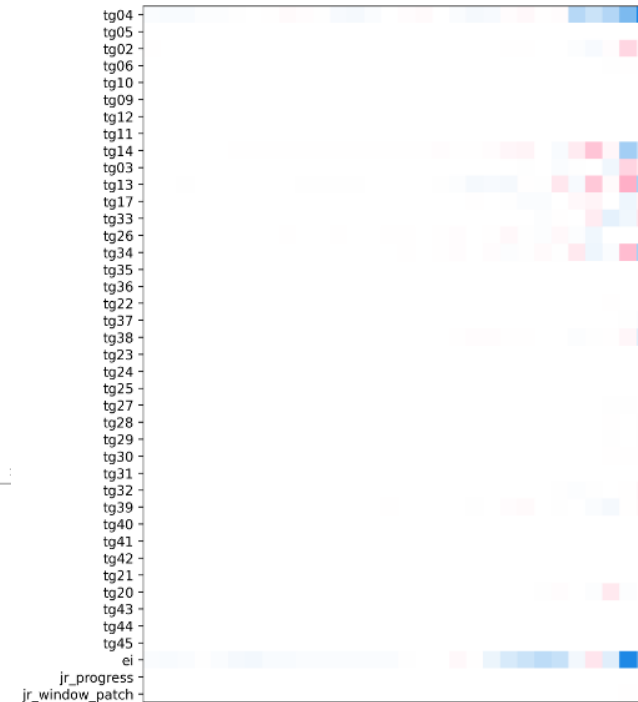
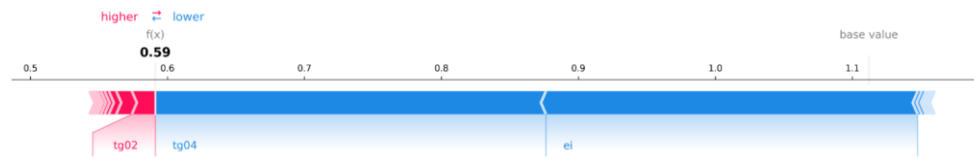


● Index : 16067

- 2023-08-23 10:35:00
- ei : 0.4201
- Type1_error_under
Predicted: 0, True: 1

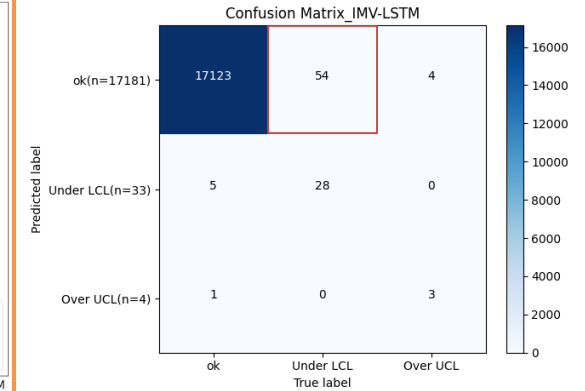
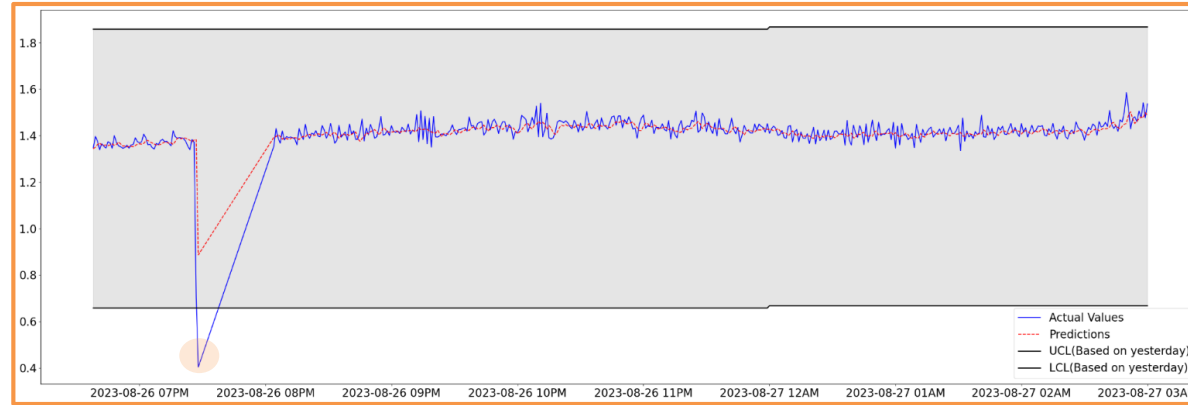


- 주요 영향 변수
tg04 : 스팀 순간값,
ei : 생산량,
tg02 : 종이별 측정 무게,
tg03 : 끝단 설비 속도,
tg20 : 스팀 압력



● Index : 16800

- 2023-08-26 19:28:00
- ei : 0.4037
- Type1_error_under
Predicted: 0, True: 1



- 주요 영향 변수
tg04 : 스팀 순간값,
tg02 : 종이별 측정 무게,
tg13 : 설비B1 절대습도,
tg38 : 설비 AE 속도,
tg20 : 스팀 압력

