

# XGboost 예측

## 수치형 컬럼

```
numeric_col = ['delivered_kwh','requested_kwh','kwh_request_diff','kwh_per_usage_time','kwh_per_usage_time_missing']
```

## 원핫 처리 컬럼

```
onehot_col =
```

```
['station_location','evse_name','evse_type','supports_discharge','scheduled_charge','weekday','usage_departure_range','post_charge_departure_range']
```

## 예측 타겟

post\_charge\_departure\_range (분류: jenkins로 5등급화)

usage\_departure\_range (분류: jenkins로 5등급화)

kwh\_per\_usage\_time (회귀: 수치값)

## 1차 결과

post\_charge\_departure\_range & usage\_departure\_range 클래스 불균형 문제 있음

kwh\_per\_usage\_time 나쁘지 않음

```
feature_names = ct.get_feature_names_out()
```

모델에서 importances 추출 결과 충전소 이름 충전기기가 대부분을 차지함

```
importances1 = pipeline_class1.named_steps['model'].feature_importances_
```

```
importances2 = pipeline_class2.named_steps['model'].feature_importances_
```

결론 : 낮음 품질의 데이터로 인해 발생하는 데이터 불균형으로 인한 분류가 불확실함

개선방향 : 분류에 불필요한 컬럼 제거후 다시 학습 진행

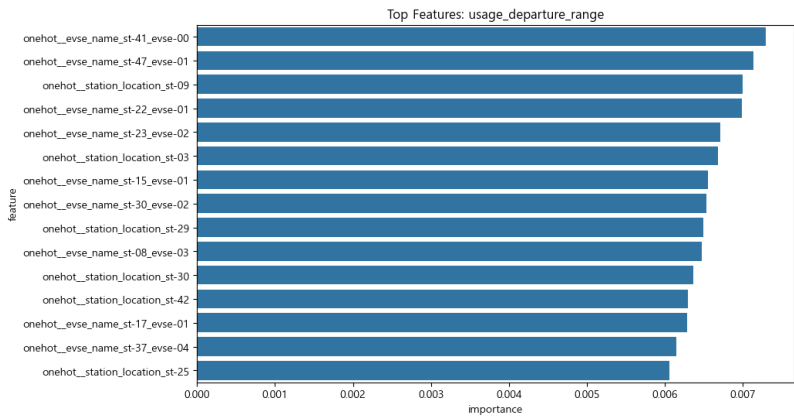
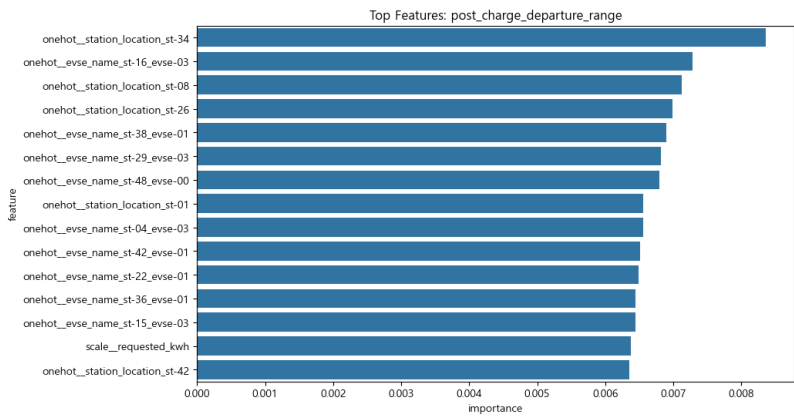
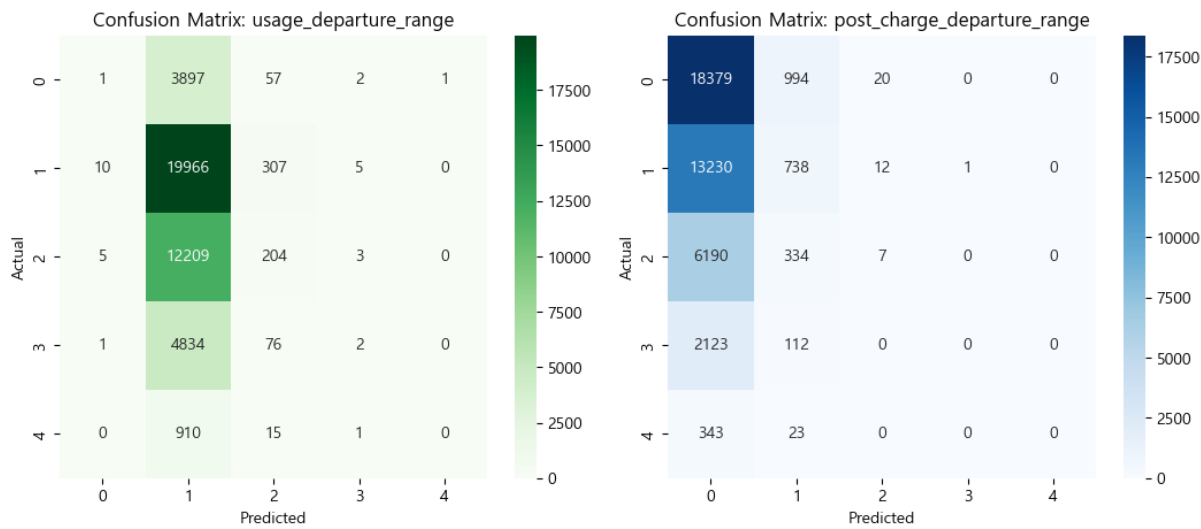
▶ post\_charge\_departure\_range: precision recall f1-score support

0	0.46	0.95	0.62	19393
1	0.34	0.05	0.09	13981
2	0.18	0.00	0.00	6531
3	0.00	0.00	0.00	2235
4	0.00	0.00	0.00	366
accuracy			0.45	42506
macro avg	0.19	0.20	0.14	42506
weighted avg	0.35	0.45	0.31	42506

▶ usage\_departure\_range: precision recall f1-score support

	0	0.06	0.00	0.00	3958
	1	0.48	0.98	0.64	20288
	2	0.31	0.02	0.03	12421
	3	0.15	0.00	0.00	4913
	4	0.00	0.00	0.00	926
accuracy				0.47	42506
macro avg	0.20	0.20	0.14		42506
weighted avg	0.34	0.47	0.32		42506

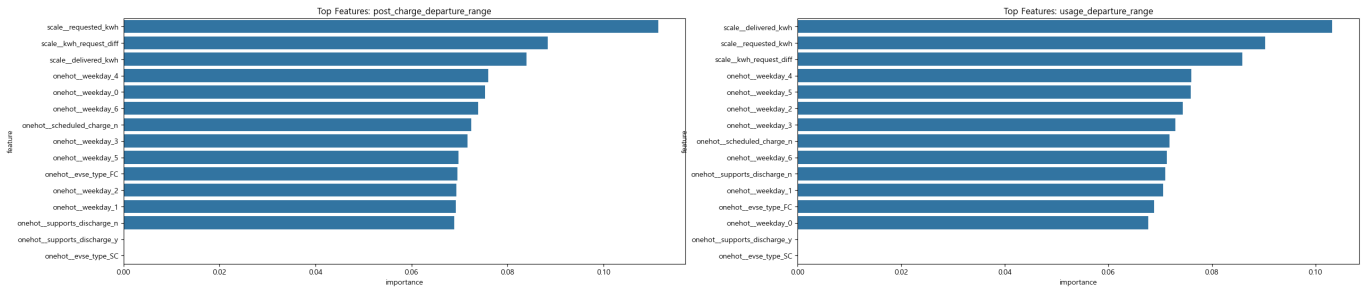
► kwh\_per\_usage\_time: MSE: 63.04574002479884 Mean, std 27.364145519869286 28.080010358688845



결론 : 충전소이름,기기명을 제거함으로 등급에 대한 분류 예측에 영향을 주는 피쳐가 다르게 분포하게 됨

충전 완료후 연결 해제 등급 분류가 2개 ->4개로 분류 할수 있게됨 하지만 여전히 예측 점수가 낮고 정확도가 낮음

개선방향 : 다음 점수를 기반으로 LSTM 모델로 예측 수행 후 비교 예정



► post\_charge\_departure\_range: precision recall f1-score support

0	0.46	0.95	0.62	19393
1	0.32	0.05	0.08	13981
2	0.21	0.00	0.00	6531
3	0.25	0.00	0.00	2235
4	0.00	0.00	0.00	366
accuracy			0.45	42506
macro avg	0.25	0.20	0.14	42506
weighted avg	0.36	0.45	0.31	42506

► usage\_departure\_range: precision recall f1-score support

0	0.00	0.00	0.00	3958
1	0.48	0.99	0.64	20288
2	0.30	0.01	0.03	12421
3	0.06	0.00	0.00	4913
4	0.00	0.00	0.00	926
accuracy			0.47	42506
macro avg	0.17	0.20	0.13	42506
weighted avg	0.32	0.47	0.31	42506

► kwh\_per\_usage\_time: MSE: 62.655199809293975 Mean, std 27.364145519869286 28.080010358688845