

Test_0825



테스트 공통 조건

- 클러스터 통일
- 변수 유지(work_time, ...)



테스트 항목 : 변수 추가

▼ 전력소비량 관련 파생변수 추가

- 추가 변수
 - 건물별, 요일별, 시간대별 평균
 - 건물별, 시간대별 평균
 - 건물별, 시간대별 표준편차



테스트 항목 : 변수 선택

▼ feature importance 0인 변수 제외

▼ 사용 변수

- Cluster 0

```
col_0 = ['weekday_hour_mean', 'month', 'sy_en', 'ma_dis_5', 'building_number', 'ess_capacity', 'work_time', 'day', 'weekda
y', 'particular', 'hour_mean', 'total_area', 'hour_std', 'CDH', 'cooling_area',
'holiday', 'discomfort_mean', 'discomfort', 'CDH_mean', 'hour_cos', 'temperature_mean', 'building_type_Research Institut
e', 'solar_power_capacity', 'humidity', 'building_type_University',
'temperature_3', 'ma_dis_3', 'hour_sin', 'pcs_capacity', 'temperature', 'low_day', 'sensory_temperature', 'discomfort_3']
```

- Cluster 1

```
col_1 = ['weekday_hour_mean', 'month', 'discomfort', 'CDH_mean', 'holiday', 'weekday', 'temperature_mean', 'hour_mean', 'build
ing_number', 'ma_dis_3', 'discomfort_mean', 'day', 'cooling_area',
'hour_std', 'total_area', 'ma_dis_5', 'CDH', 'humidity', 'sensory_temperature', 'hour_sin', 'temperature', 'temperature_3', 'h
our_cos', 'work_time', 'discomfort_3']
```

- Cluster 2

```
col_2 = ['weekday_hour_mean', 'building_number', 'hour_mean', 'month', 'ma_dis_5', 'cooling_area', 'discomfort', 'CDH', 'da
y', 'hour_std', 'hour_sin', 'total_area', 'humidity', 'holiday', 'hour_cos', 'CDH_mean',
'weekday', 'ma_dis_3', 'work_time', 'discomfort_mean', 'temperature_mean', 'temperature_3', 'sensory_temperature']
```

- Cluster 3

```
col_3 = ['weekday_hour_mean', 'month', 'hour_std', 'ma_dis_5', 'holiday', 'discomfort', 'weekday', 'building_type_Commercia
l', 'low_day', 'day', 'CDH', 'total_area', 'building_type_Hospital', 'hour_mean',
'building_number', 'CDH_mean', 'solar_power_capacity', 'ma_dis_3', 'hour_sin', 'cooling_area', 'discomfort_mean', 'humidit
y', 'sensory_temperature', 'hour_cos', 'temperature_mean', 'sy_en',
'windspeed', 'work_time', 'ess_capacity']
```

▼ 테스트 : Pycaret - et 지정 모델

- Cluster 0
정규화 수정 전 - SMAPE : 4.08768
정규화 수정 후 - SMAPE : 4.08667 + sin, cos 수정 후 - SMAPE : 3.92282
- Cluster 1
정규화 수정 전 - SMAPE : 4.67567
정규화 수정 후 - SMAPE : 4.67014 + sin, cos 수정 후 - SMAPE : 4.63359
- Cluster 2
정규화 수정 전 - SMAPE : 2.31264
정규화 수정 후 - SMAPE : 2.29887 + sin, cos 수정 후 - SMAPE : 2.15511
- Cluster 3
정규화 수정 전 - SMAPE : 2.93476
정규화 수정 후 - SMAPE : 2.93204 + sin, cos 수정 후 - SMAPE : 2.86132

▼ 테스트 : Pycaret - compare model

- Cluster 0 : et
정규화 수정 전 - SMAPE : 4.09256
정규화 수정 후 - SMAPE : 4.09870 +sin, cos 수정 후 - SMAPE : 3.91646
- Cluster 1 : et
정규화 수정 전 - SMAPE : 4.69697
정규화 수정 후 - SMAPE : 4.71004 + sin, cos 수정 후 - SMAPE : 4.65683
- Cluster 2 : Catboost
정규화 수정 전 - SMAPE : 2.19484
정규화 수정 후 - SMAPE : 2.20908 + sin, cos 수정 후 - SMAPE : 2.23476
- Cluster 3 : et
정규화 수정 전 - SMAPE : 2.94628
정규화 수정 후 - SMAPE : 2.93667 + sin, cos 수정 후 - SMAPE : 2.84904

▼ Shap 결과 전부 사용

▼ 사용 변수

- Cluster 0

```
['weekday_hour_mean', 'month', 'ma_dis_5', 'CDH', 'day', 'discomfort', 'building_number', 'holiday', 'weekday', 'hour_std', 'discomfort_mean', 'hour_mean', 'total_area', 'CDH_mean', 'humidity', 'cooling_area', 'ma_dis_3']
```

- Cluster 1

```
['weekday_hour_mean', 'month', 'discomfort', 'day', 'temperature_3', 'CDH', 'hour_std', 'discomfort_mean', 'discomfort_3', 'hour_mean', 'weekday', 'temperature', 'ma_dis_3', 'ma_dis_5', 'temperature_mean', 'building_number', 'humidity', 'cooling_area']
```

- Cluster 2

```
['weekday_hour_mean', 'building_number', 'hour_mean', 'month', 'ma_dis_5', 'CDH', 'discomfort', 'day', 'hour_sin', 'humidity', 'hour_std', 'cooling_area', 'weekday', 'temperature_3', 'CDH_mean', 'ma_dis_3', 'hour_cos', 'holiday', 'discomfort_mean']
```

- Cluster 3

```
['weekday_hoyur_mean', 'month', 'holiday', 'discomfort', 'ma_dis_5', 'day', 'CDH', 'hour_std', 'hour_sin', 'weekday', 'hour_mean', 'ma_dis_3', 'CDH_mean', 'total_area', 'building_number', 'humidity', 'cooling_area', 'discomfort_mean', 'sensory_temperature']
```

▼ 테스트 : Pycaret - et 지정 모델

- Cluster 0

SMAPE : 4.08973

sin, cos 수정 후 : SMAPE : 3.96421

- Cluster 1

SMAPE : 4.66345

sin, cos 수정 후 : SMAPE : 4.66630

- Cluster 2

SMAPE : 2.27170

sin, cos 수정 후 : SMAPE : 2.14489

- Cluster 3

SMAPE : 2.94844

sin, cos 수정 후 : SMAPE : 2.87411

▼ 테스트 : Pycaret - compare model

- Cluster 0 : et

SMAPE : 4.07536

sin, cos 수정 후 : SMAPE : 3.98163

- Cluster 1 : et

SMAPE : 4.65229

sin, cos 수정 후 : SMAPE : 4.68179

- Cluster 2 : CatBoost

SMAPE : 2.12017

sin, cos 수정 후 : SMAPE : 2.24154

- Cluster 3

SMAPE : 2.95855

sin, cos 수정 후 : SMAPE : 2.84979

▼ 변수 중요도 0 제외, 하위 2개 제거

▼ 사용 변수

- Cluster 0

```
['weekday_hour_mean', 'date_sin', 'CDH', 'temperature_mean', 'building_number',  
'hour_std', 'hour_mean', 'discomfort_mean', 'discomfort', 'holiday', 'CDH_mean',  
'ma_dis_3', 'hour_sin', 'total_area', 'cooling_area', 'cooling_area', 'weekday_sin',  
'weekday_cos', 'ma_dis_5', 'solar_power_capacity']
```

- Cluster 1

```
['weekday_hour_mean', 'date_sin', 'discomfort', 'temperature_mean', 'discomfort_mean',  
'hour_std', 'CDH', 'building_number', 'humidity', 'total_area', 'ma_dis_5', 'temperature',  
'hour_sin', 'weekday_cos', 'ma_dis_3']
```

- Cluster 2

```
['weekday_hour_mean', 'hour_mean', 'building_number', 'date_sin', 'ma_dis_5',  
'discomfort', 'CDH', 'hour_sin', 'hour_std', 'humidity', 'total_area']
```

- Cluster 3

```
['weekday_hour_mean', 'date_sin', 'holiday', 'hour_mean', 'hour_std', 'CDH',  
'sensory_temperature', 'discomfort', 'hour_sin', 'weekday_sin', 'temperature_mean',
```

```
'ma_dis_5', 'total_area', 'building_number', 'CDH_mean', 'discomfort_mean']
```

▼ 테스트 : Pycaret - et 지정 모델

- Cluster 0
SMAPE : 4.05002
- Cluster 1
SMAPE : 4.73512
- Cluster 2
SMAPE : 2.43135
- Cluster 3
SMAPE : 2.92323

▼ 테스트 : Pycaret - compare model

- Cluster 0 : et
SMAPE : 4.05499
- Cluster 1 : et
SMAPE : 4.69945
- Cluster 2 : CatBoost
SMAPE : 2.27453
- Cluster 3 : et
SMAPE : 2.92448

▼ 변수 중요도 0 제외, 하위 5개 제거

▼ 상관관계 높은 변수 제거



테스트 항목 : 변수 처리

*변수 선택 테스트 결과를 반영하지 않은 것으로, 상대적 비교만 할 것

*et 단일 모델 기준

▼ Flag 변수 : boolean vs int

- Cluster 0
 - Bool : SMAPE : 4.06987
 - int : SMAPE : 4.07696
- Cluster 1
 - Bool : SMAPE : 4.67106
 - int : SMAPE : 4.67076
- Cluster 2
 - Bool : SMAPE : 2.28308
 - int : SMAPE : 2.27471
- Cluster 3
 - Bool : SMAPE : 2.94351
 - int : SMAPE : 2.95340

▼ 로그 변환 유무

- Cluster 0
 - 로그 변환 O : SMAPE : 4.08192

- 로그 변환 X : SMAPE : 4.14520
- Cluster 1
 - 로그 변환 O : SMAPE : 4.67336
 - 로그 변환 X : SMAPE : 4.63916
- Cluster 2
 - 로그 변환 O : SMAPE : 2.27023
 - 로그 변환 X : SMAPE : 2.29059
- Cluster 3
 - 로그 변환 O : SMAPE : 2.96356
 - 로그 변환 X : SMAPE : 2.99653
- ▼ standardscaling 유무
 - Cluster 0
 - standardscaling O : SMAPE : 4.05704
 - standardscaling X : SMAPE : 4.07858
 - Cluster 1
 - standardscaling O : SMAPE : 4.66351
 - standardscaling X : SMAPE : 4.67746
 - Cluster 2
 - standardscaling O : SMAPE : 2.26480
 - standardscaling X : SMAPE : 2.27325
 - Cluster 3
 - standardscaling O : SMAPE : 2.95955
 - standardscaling X : SMAPE : 2.96328