



# 2025 전력사용량 예측 AI 경진대회

서동현, 유동현



# 목차

Dacon Contents

## 배경

평가 지표 및 핵심 전략

01p

## 데이터

대회 데이터 및 데이터셋 구성  
EDA  
데이터 전처리  
피쳐 엔지니어링

03p

## 모델링

Xgboost  
Lightgbm  
Catboost  
Comparison  
Staking

11p

## 최종 제출물

Ensemble

21p

# 배경

Background

## 건물의 전력사용량 예측 AI 모델 개발

안정적·효율적인 에너지 공급을 위해 전력 사용량 예측 중요성 확대

기후 변화 및 에너지 전환 정책에 따른 수요 예측 기반 관리 역량 필요

2025 전력사용량 예측 AI 경진대회 개최 (2021·2023 이어 세 번째)

산업 현장에 활용 가능한 전력 수요 예측 알고리즘 발굴 및 에너지 분야 AI 기술 적용 가능성 모색

## 2025 전력사용량 예측 AI 경진대회

알고리즘 | 정형 | 시계열 | 에너지 | SMAPE

₩ 상금 : 2,000 만원

🕒 2025.07.14 ~ 2025.08.25 09:59

[+ Google Calendar](#)

👤 1,526명 📅 마감

# 평가 지표 및 핵심 전략

Evaluation Metrics & Key Strategies

## 평가 지표

**SMAPE** (Symmetric Mean Absolute Percentage Error)

$$SMAPE = \frac{100}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

## 핵심 전략

**Feature Engineering** : 기상학적, 공학적, 시계열 피쳐 생성

**Modeling** : xgboost, catboost, lgbm

**Ensemble** : 시드, fold ensemble, 모델별 최종 파일 Ensemble

건물별 개별 모델링 (100개 건물)

유형별 개별 모델

# 대회 데이터 및 데이터셋 구성

About Data

## 입력데이터

건물 특성 : 건물 유형, 연면적, 냉방면적, 태양광, ESS저장용량, PCS용량

기상 정보 : 기온, 강수량, 풍속, 습도, 일조, 일사량

## 출력 데이터

전력 사용량 예측 : 특정 시점의 건물별 전력 사용량(kWh)

예측 기간 : 2024.08.25 ~ 08.31 (7일간)

파일명	내용
train.csv	2024.06.01~08.24 100개 건물의 기상 데이터, 실제 전력사용량
building_info.csv	건물 번호, 건물 유형별 연면적, 냉방면적, 태양광 ESS, PCS 용량
test.csv	2024.08.25 ~ 08.31 100개 건물의 기상 데이터
sample_submission.csv	제출 양식 (ID: 건물번호 + 시간 / answer: 예측값)

# 데이터

Trian Dataset

Train Dataset

num_date_time	건물번호	일시	기온(°C)	강수량(mm)	풍속(m/s)	습도(%)	일조(hr)	일사(MJ/m2)	전력소비량(kWh)
1_2024060100	1	2024060100	18.3	0	2.6	82	0	0	5794.8
1_2024060101	1	2024060101	18.3	0	2.7	82	0	0	5591.85
1_2024060102	1	2024060102	18.1	0	2.6	80	0	0	5338.17
1_2024060103	1	2024060103	18	0	2.6	81	0	0	4554.42
1_2024060104	1	2024060104	17.8	0	1.3	81	0	0	3602.25
1_2024060105	1	2024060105	17.7	0	2.2	83	0	0	3367.59
1_2024060106	1	2024060106	17.6	0	1.7	85	0	0.02	3089.4
1_2024060107	1	2024060107	17.7	0	1.2	81	0	0.22	3977.67
1_2024060108	1	2024060108	18.2	0	1.8	78	0	0.56	4380.45
1_2024060109	1	2024060109	19.4	0	2.4	70	0.4	1.27	4359.78
1_2024060110	1	2024060110	20.5	0	2.7	65	0.7	2.36	4494.72
1_2024060111	1	2024060111	21.7	0	3.4	60	0.8	3.02	4566.75
1_2024060112	1	2024060112	22.7	0	3.1	53	0.9	3.4	4591.29
1_2024060113	1	2024060113	23.1	0	3.5	48	0.8	3.29	4596.6

Target

Building\_info Dataset

	건물유형	연면적(m2)	냉방면적(m2)	태양광용량(kW)	ESS저장용량(kWh)	PCS용량(kW)
1	호텔	82912.71	77586	-	-	-
2	상용	40658.9	30392.82	-	-	-
3	병원	560431	418992	278.58	-	-
4	호텔	41813.29	23715.71	-	-	-
5	학교	403749.39	248507	1983.05	1025	250
6	상용	157835	157835	-	-	-
7	건물기타	118346	78237.38	389.76	-	-
8	학교	167751	82112	217.92	-	-
9	호텔	136757	109024	-	-	-
10	호텔	435993.5	341983	-	-	-
11	아파트	271233	233263	-	-	-
12	학교	581897	311452	1349.03	-	-
13	연구소	97372.12	44536	276	-	-
14	학교	504332	339131	849.78	101.5	101

Join

# 데이터

Test Dataset

num_date_time	건물번호	일시	기온(°C)	강수량(mm)	풍속(m/s)	습도(%)
1_20240825 00	1	20240825 00	26.5	0	0.7	80
1_20240825 01	1	20240825 01	26.1	0	0	80
1_20240825 02	1	20240825 02	25.9	0	0.3	83
1_20240825 03	1	20240825 03	25.7	0	1.1	83
1_20240825 04	1	20240825 04	25.5	0	1	86
1_20240825 05	1	20240825 05	25.2	0	1.4	88
1_20240825 06	1	20240825 06	25	0	1.7	90
1_20240825 07	1	20240825 07	25.1	0	1.6	90
1_20240825 08	1	20240825 08	26.1	0	1.5	89
1_20240825 09	1	20240825 09	27.4	0	1.5	84
1_20240825 10	1	20240825 10	28.5	0	1.2	77
1_20240825 11	1	20240825 11	29.4	0	0.7	74
1_20240825 12	1	20240825 12	30.7	0	1.5	69
1_20240825 13	1	20240825 13	31.5	0	1.5	64



추가  
피쳐엔지니어링

예측

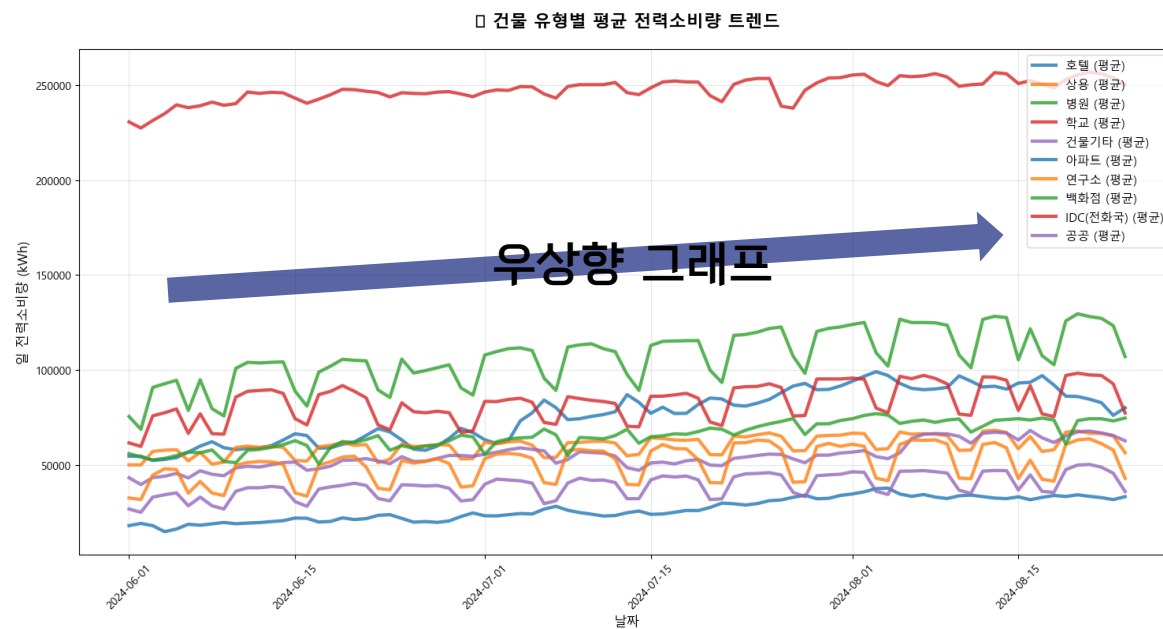
제출 답안

num_date_time	answer
1_20240825 00	0
1_20240825 01	0
1_20240825 02	0
1_20240825 03	0
1_20240825 04	0
1_20240825 05	0
1_20240825 06	0
1_20240825 07	0
1_20240825 08	0
1_20240825 09	0
1_20240825 10	0
1_20240825 11	0
1_20240825 12	0
1_20240825 13	0

# EDA

## Exploratory Data Analysis

건물 유형	건물 수	평균 소비량 (kWh)
호텔	10개	3175.02
상용	10개	2513.70
병원	9개	4454.06
학교	10개	3462.68
건물기타	10개	2285.97
아파트	9개	1106.31
연구소	9개	2111.67
백화점	16개	2729.74
IDC(전화국)	9개	10316.94
공공	8개	1625.92
합계	100개	-

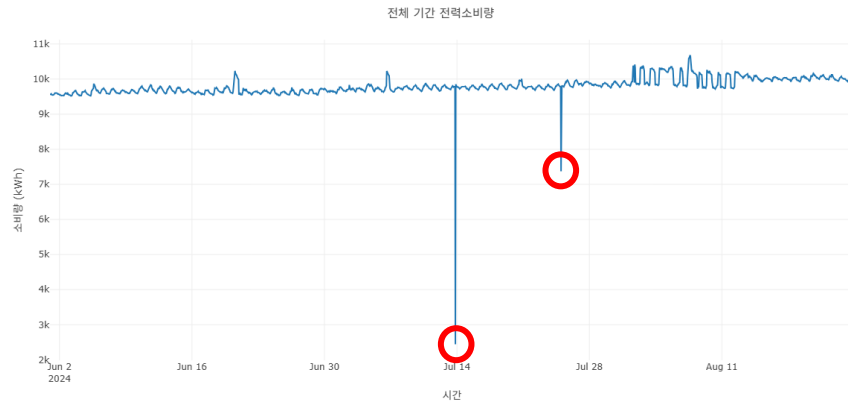


유형별 전력 소비량(Target)의 패턴 존재  
 여름철 냉방 수요로 인한 전반적인 우상향 그래프  
 소비 패턴의 차이로 인한 유형별, 건물별 모델링 전략 수립



# EDA

## Exploratory Data Analysis

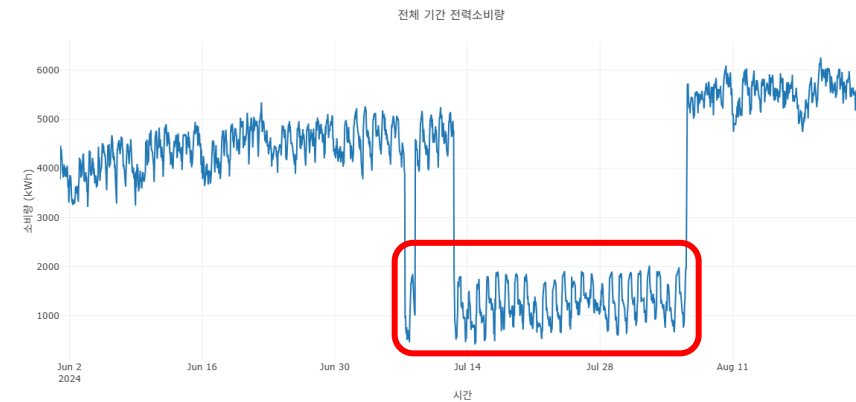


### 이상치 존재 건물:

건물별 패턴 분석 결과 일부 건물들의 경우 전체적인 추세와 다른 이상치 존재  
향후 처리 전략 수립

### 불규칙 패턴 존재 건물:

전체적인 패턴이 존재안에서 불규칙하게 나타나는 패턴  
예측 기간의 데이터에서 이러한 패턴도 예측해야한다고 판단 (유지)



# 데이터 전처리

## Data Preprocessing

### 데이터 전처리 및 보간

#### 향상된 스플라인 보간 (Enhanced Spline Interpolation)

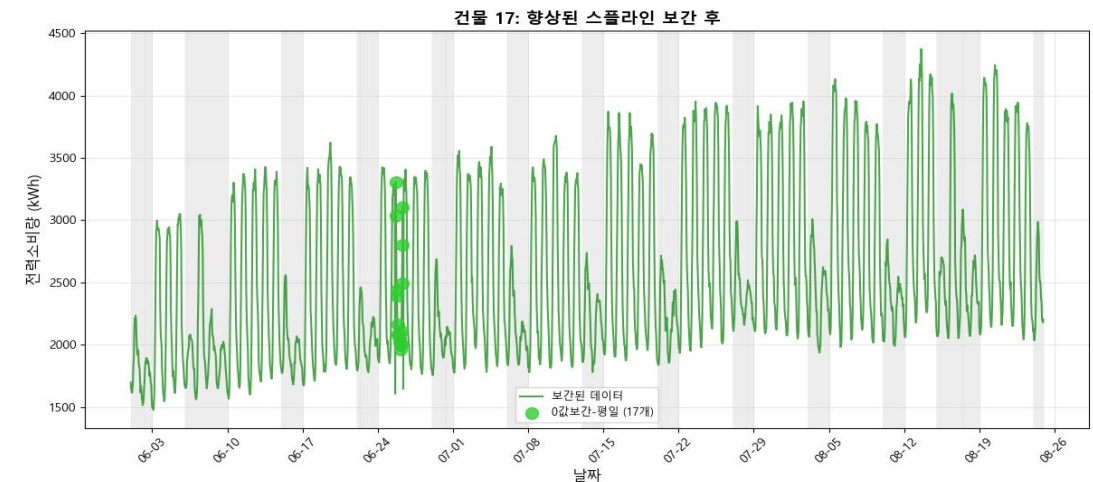
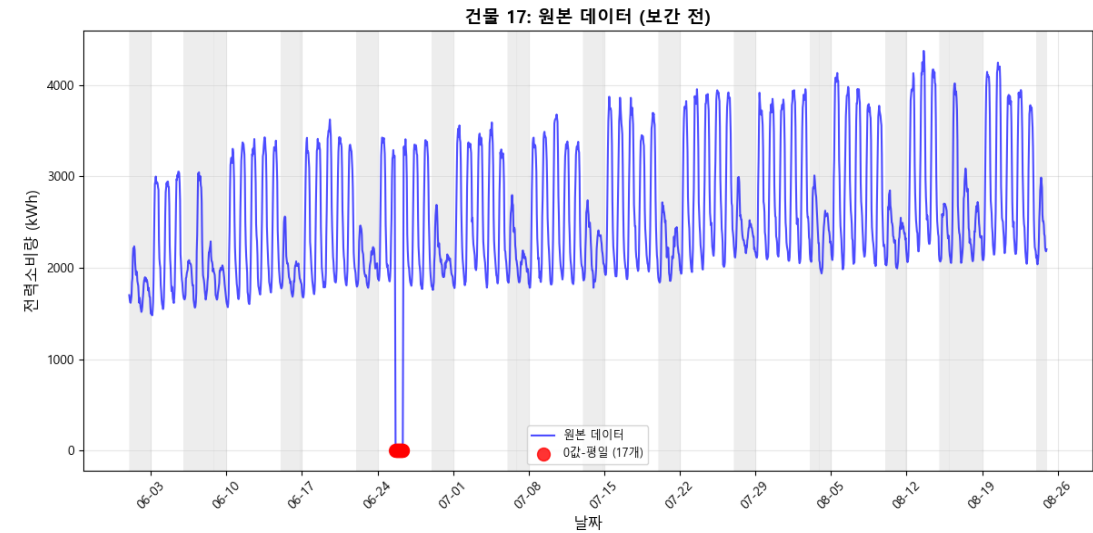
공식 :  $S(w) = a_i + b_i(w - w_i) + c_i(w - w_i)^2 + d_i(w - w_i)^3$

**0값 보간:** 전력소비량이 0인 값들을 패턴 기반으로 보간

**이상치 보간:** 특정 건물별 이상치 패턴을 정의하여 보간

**방법:** CubicSpline을 사용한 주차 단위 패턴 매칭

**패턴 매칭:** 같은 시간대, 같은 요일, 같은 휴일 여부를 고려



# 피쳐 엔지니어링

## Feature Engineering

피쳐명	카테고리	의미/설명
sin_hour	시간 순환	시간 사인 인코딩
cos_hour	시간 순환	시간 코사인 인코딩
sin_date	시간 순환	월-일 사인 인코딩
cos_date	시간 순환	월-일 코사인 인코딩
sin_month	시간 순환	월 사인 인코딩
cos_month	시간 순환	월 코사인 인코딩
sin_dayofweek	시간 순환	요일 사인 인코딩
cos_dayofweek	시간 순환	요일 코사인 인코딩
holiday	특별 패턴	건물별 휴무 여부
weather	특별 패턴	날씨 영향(강수 전후)
summer_cos	계절 패턴	여름철 코사인 패턴
summer_sin	계절 패턴	여름철 사인 패턴
week_of_month	특별 패턴	월내 격주 일요일 패턴

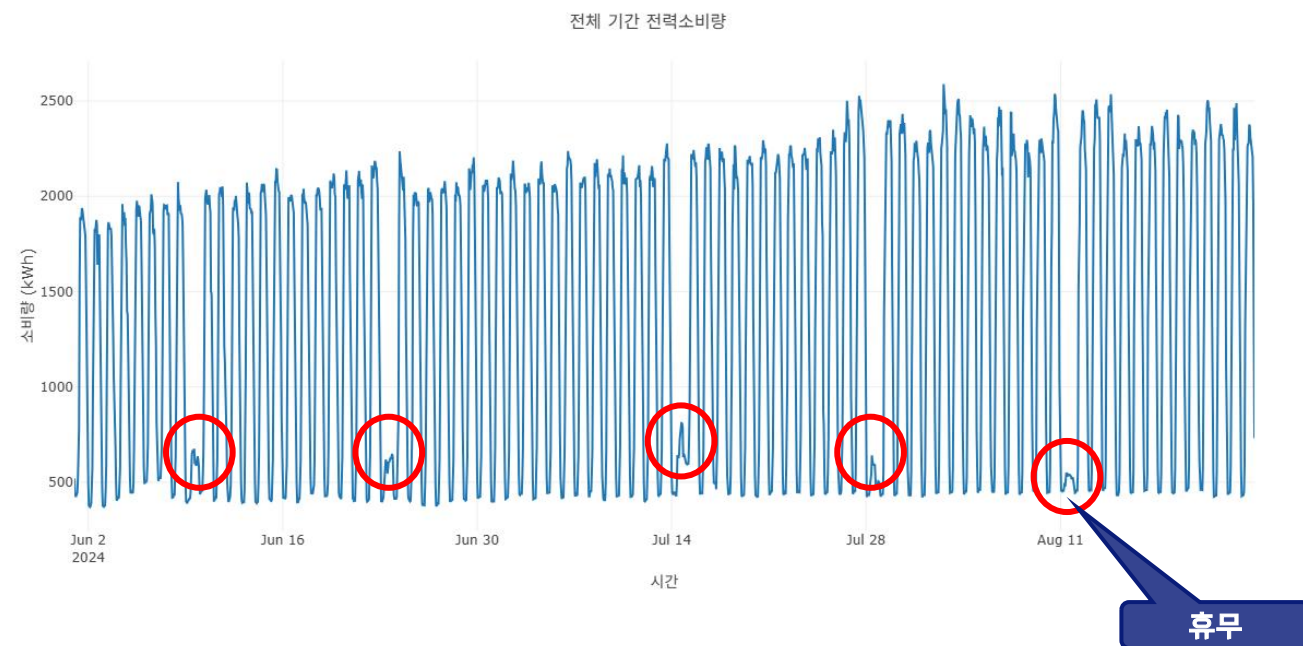
피쳐명	카테고리	의미/설명
enhanced_avg_temp	온도 통계	3시간 간격 일평균 온도
day_mean_temperature	온도 통계	일평균 온도
day_temperature_range	온도 통계	일교차
CDH	복합 지수	냉방도(12시간 누적)
THI	복합 지수	온도-습도 지수
WCT	복합 지수	체감온도
day_hour_mean	전력 통계	요일-시간별 과거 평균
day_hour_std	전력 통계	요일-시간별 과거 표준편차
holiday_hour_mean	전력 통계	휴일-시간별 과거 평균
holiday_hour_std	전력 통계	휴일-시간별 과거 표준편차
hour_mean	전력 통계	시간별 과거 평균
hour_std	전력 통계	시간별 과거 표준편차
hot_humid_condition	환경 조건	고온다습 조건

# 피쳐 엔지니어링

Feature Engineering

## 백화점 건물별 특별 휴무 패턴

피처명	휴무 패턴
building_18	매주 일요일 휴무
building_19	특정 날짜 휴무
building_27	격주 일요일 휴무
building_32	매월 10일 + 특정 일요일 휴무
building_34	격주 월요일 휴무
building_40	격주 일요일 휴무 (building_27과 동일)
building_45	특정 날짜 휴무 (building_19와 동일)
building_54	격주 일요일 휴무 (building_27과 동일)
building_59	격주 일요일 휴무 (building_27과 동일)
building_63	격주 일요일 휴무 (building_27과 동일)
building_73	연중무휴
building_74	복합 휴무 (격주 일요일 + 추가 월요일)
building_79	특정 날짜 휴무 (building_54와 동일)
building_88	연중무휴
building_95	특정 날짜 휴무
기타 건물들	주말 + 공휴일 휴무 (일반 패턴)



# 모델\_Xgboost

Model\_XGboost

## 단일 모델

모델: XGBoost Regressor (그래디언트 부스팅)

### 모델 파라미터

learning\_rate: 0.1,  
n\_estimators: 1000,  
max\_depth: 5  
subsample: 0.9,  
colsample\_bytree: 0.8  
min\_child\_weight: 6,  
reg\_alpha: 1,  
reg\_lambda: 1

### 단계별 가중치 개선

Weighted MSE: 언더예측에 100배 가중치

요일별 성능 분석: 월~일요일별 SMAPE 측정

## 2단계 앙상블 전략

1단계: 각 폴드별 시드 앙상블

• 5개 시드 [2025, 42, 123, 777, 999]

• 폴드당 시드별 예측 평균

2단계: 폴드 앙상블

• 12-Fold 교차검증 (7일 단위 역순)

• validation 앞뒤 1일 drop으로 데이터 리키지 방지

### 성능 결과

검증 방식: 시계열 12-Fold 교차검증

최종 앙상블: 시드 × 폴드 = 60개 모델 평균

## 건물별/유형별 분할 전략

건물별 모델 : LB Score 5.4208

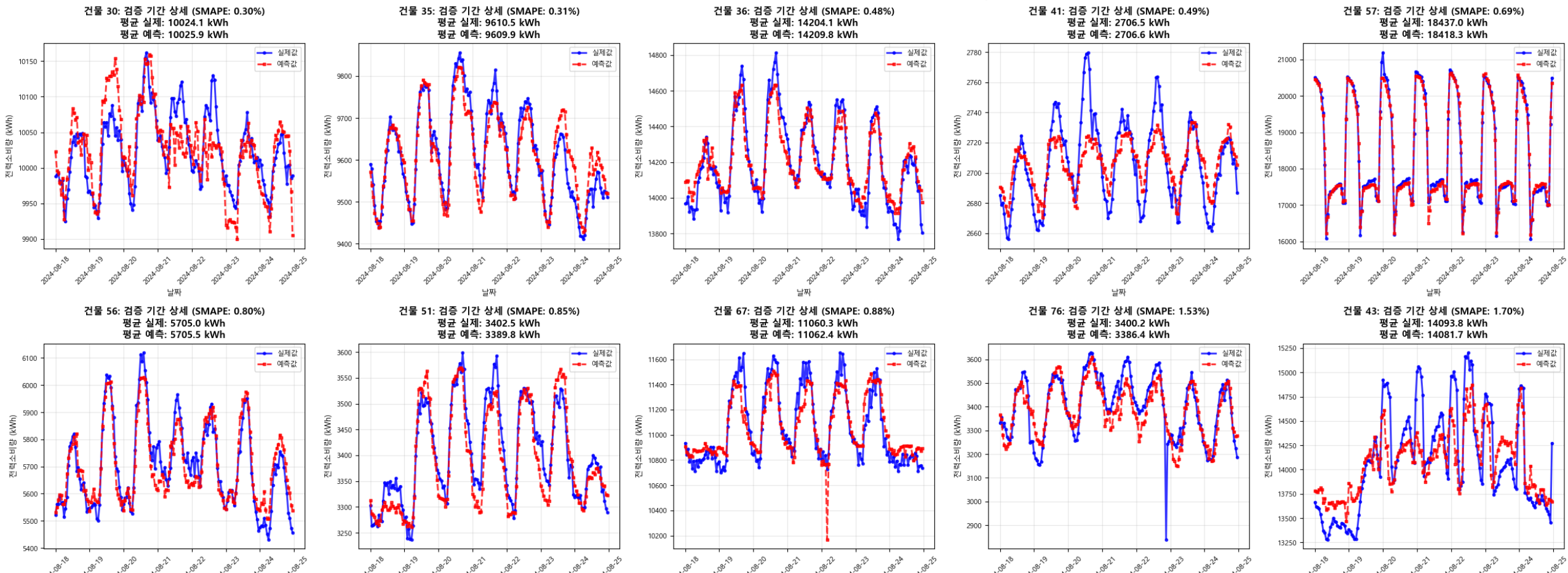
유형별 모델 : LB Score 5.7983

# 모델\_Xgboost

Model\_XGboost

트리기반 모델/ 과소추정으로 예측

## XGBoost 상위 10개 건물 검증 결과 (전체 평균 SMAPE: 5.73%)



# 모델\_LightGBM

Model\_Lightgbm

## 단일 모델

모델: LightGBM Regressor (그래디언트 부스팅)

모델 파라미터

learning\_rate: 0.05,

n\_estimators: 2000,

max\_depth: 5

num\_leaves: 50,

subsample: 0.9,

colsample\_bytree: 0.8

min\_child\_samples: 15,

reg\_alpha: 1,

reg\_lambda: 1

단계별 가중치 개선

1단계: 초기 모델로 과소추정 패턴 탐지

2단계: 가중치 적용 모델 ( $\alpha=1$ )

## 2단계 앙상블 전략

폴드 앙상블

• 12-Fold 교차검증 (7일 단위 역순)

• validation 앞뒤 1일 drop으로 데이터 리키지 방지

• 최종 폴드별 예측 평균

성능 결과

검증 방식: 시계열 12-Fold 교차검증

최종 앙상블: 폴드 = 12개 모델 평균

건물별 개별: 100개 건물 각각 모델링

## 최종 생성 모델

건물별 모델 : LB Score 5.3909

유형별 모델 : LB Score 5.8168

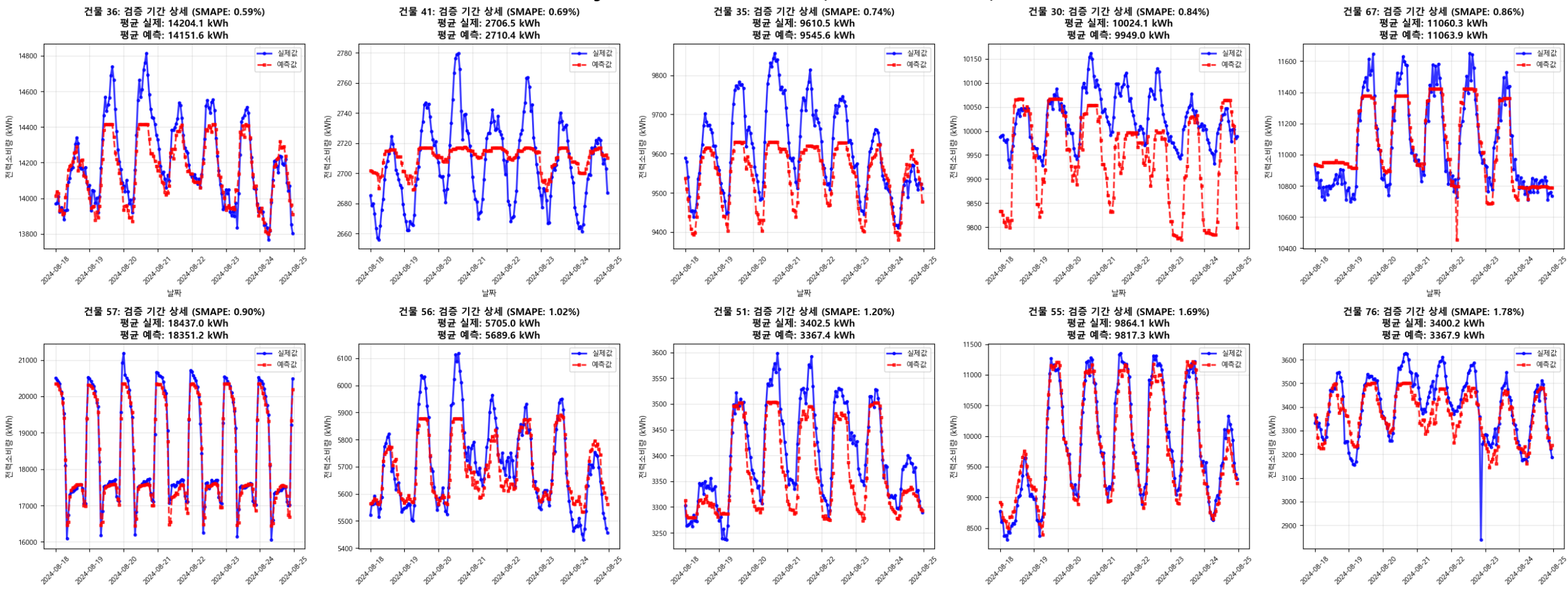


# 모델\_LightGBM

Model\_Lightgbm

단일 모델 성능은 높지 않으나, 추후 앙상블의 일반화 성능 향상 가능성

LightGBM 상위 10개 건물 검증 결과 (전체 평균 SMAPE: 5.90%)





# 모델\_Catboost

Model\_Catboost

## 단일 모델

모델: CatBoost Regressor (트리 기반)

모델 파라미터

learning\_rate: 0.1,

iterations: 1000,

depth: 5

subsample: 0.9,

colsample\_bylevel: 0.8

로그 변환:  $\log(y + 1e-6)$ , 음수 방지 적용

단계별 가중치 개선

1단계: 초기 모델로 과소추정 패턴 탐지

2단계: 가중치 적용 모델 ( $\alpha=1$ )

## 2단계 앙상블 전략

1단계: 각 폴드별 시드 앙상블

• 5개 시드 [2025, 42, 123, 777, 999]

• 폴드당 시드별 예측 평균

2단계: 폴드 앙상블

• 12-Fold 교차검증 (7일 단위 역순)

• validation 앞뒤 1일 drop으로 데이터 리키지 방지

성능 결과

검증 방식: 시계열 12-Fold 교차검증

최종 앙상블: 시드  $\times$  폴드 = 60개 모델 평균

## 건물별/유형별 분할 전략

건물별 모델 : LB Score 5.4208

유형별 모델 : LB Score 5.7983

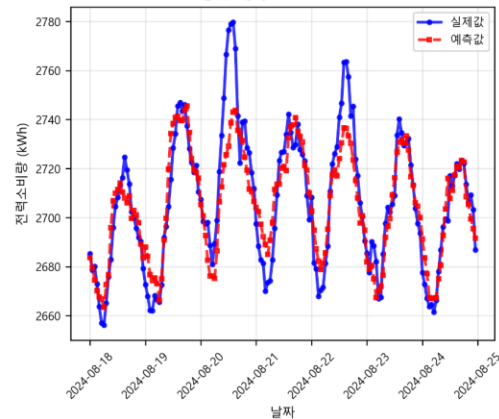
# 모델\_Catboost

Model\_Catboost

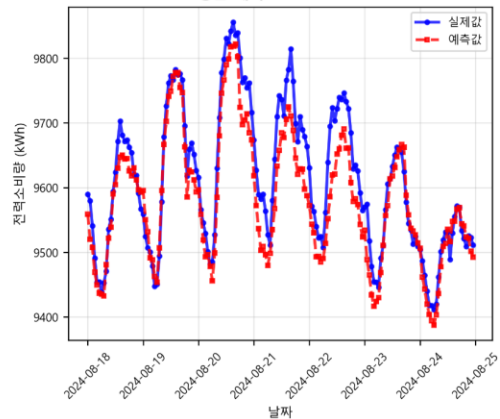
단일 모델로는 최고 성능/ 추가 보정없이 과대추정

상위 10개 건물 검증 결과 (전체 평균 SMAPE: 5.45%)

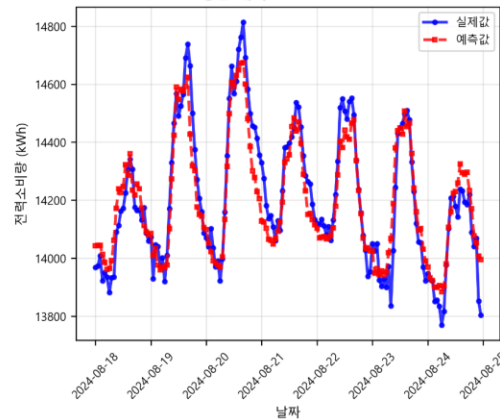
건물 41: 검증 기간 상세 (SMAPE: 0.32%)  
평균 실제: 2706.5 kWh  
평균 예측: 2705.1 kWh



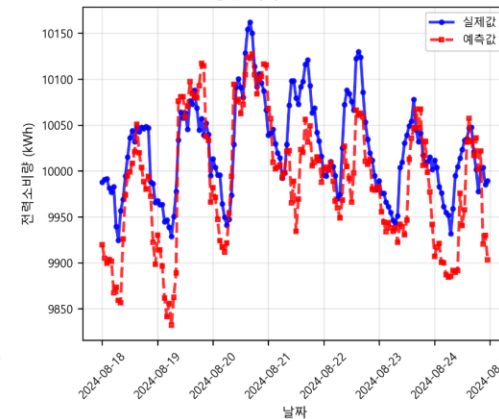
건물 35: 검증 기간 상세 (SMAPE: 0.35%)  
평균 실제: 9610.5 kWh  
평균 예측: 9581.4 kWh



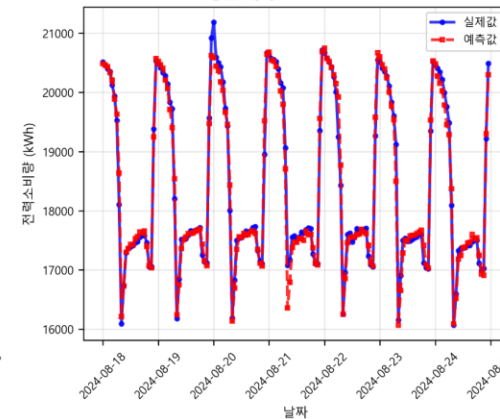
건물 36: 검증 기간 상세 (SMAPE: 0.43%)  
평균 실제: 14204.1 kWh  
평균 예측: 14197.2 kWh



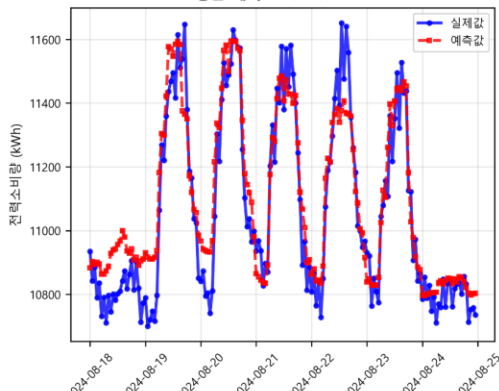
건물 30: 검증 기간 상세 (SMAPE: 0.44%)  
평균 실제: 10024.1 kWh  
평균 예측: 9989.3 kWh



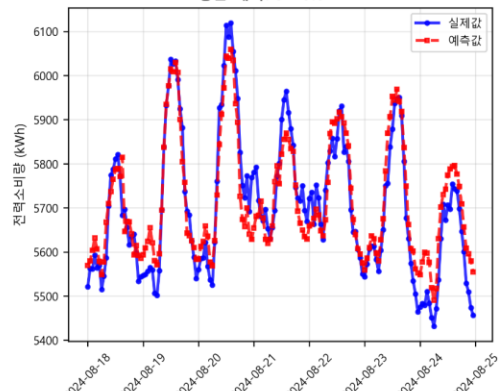
건물 57: 검증 기간 상세 (SMAPE: 0.60%)  
평균 실제: 18437.0 kWh  
평균 예측: 18418.3 kWh



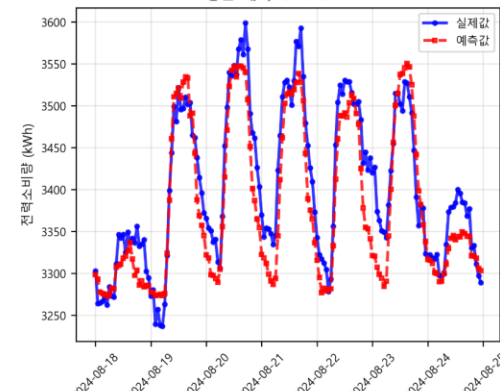
건물 67: 검증 기간 상세 (SMAPE: 0.69%)  
평균 실제: 11060.3 kWh  
평균 예측: 11096.6 kWh



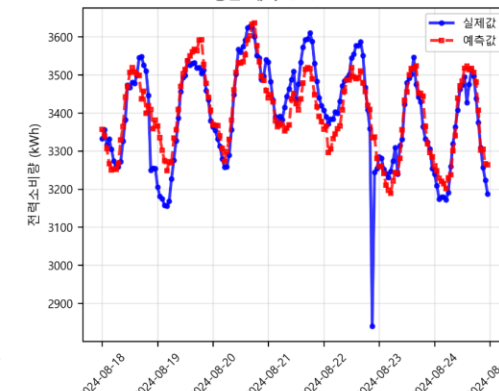
건물 56: 검증 기간 상세 (SMAPE: 0.81%)  
평균 실제: 5705.0 kWh  
평균 예측: 5715.8 kWh



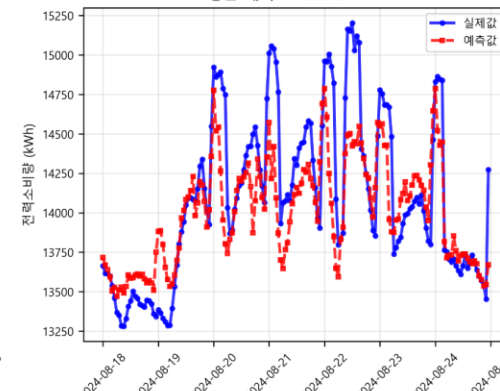
건물 51: 검증 기간 상세 (SMAPE: 0.91%)  
평균 실제: 3402.5 kWh  
평균 예측: 3381.3 kWh



건물 76: 검증 기간 상세 (SMAPE: 1.36%)  
평균 실제: 3400.2 kWh  
평균 예측: 3402.4 kWh



건물 43: 검증 기간 상세 (SMAPE: 1.59%)  
평균 실제: 14093.8 kWh  
평균 예측: 14006.5 kWh



# 모델\_Comparison

Model\_Comparison

LB Score: 5.22

## 건물별 최고 성능 모델 선택

모델: Xgboost, LightGBM, Catboost

모델 파라미터

앞서 진행한 단일 모델의 최고 성능 파라미터 사용

로그 변환:  $\log(y + 1e-6)$ , 음수 방지 적용

건물별 최고 성능 모델 선택

각 건물별 3개의 모델을 학습 및 예측 진행

각 모델별 SMAPE가 가장 낮게 나온 모델을 해당 건물의 모델로 선택

building_number	best_model	Xgboost_smape	Lightgbm_smape	Catboost_smape
1	Catboost	10.66259	10.6841	10.10088
2	Lightgbm	5.772093	4.995961	5.666383
3	Catboost	2.984743	4.259478	2.844345
⋮				
98	Xgboost	9.265551	10.21471	12.00052
99	Xgboost	3.136065	3.868694	3.759933
100	Catboost	15.03311	14.80285	13.23687

Xgboost  
LightGBM  
Catboost



각 건물별  
SMAPE 비교

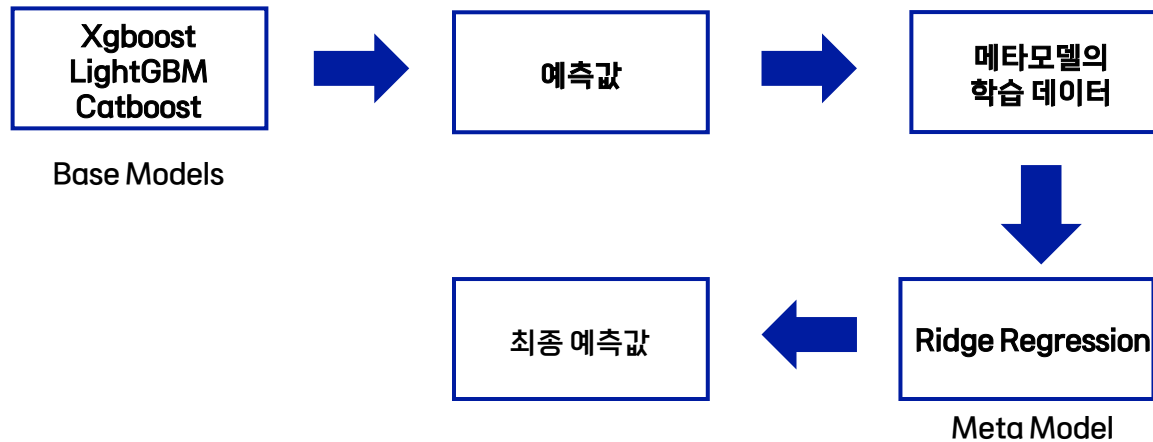


SMAPE가 가장  
좋은 모델의  
예측값 선택

# 모델\_Stacking

Model\_Stacking

LB Score: 5.2633



## Base Models

- XGBRegressor (28개 피쳐)

파라미터: learning\_rate=0.1, max\_depth=5, subsample=0.9, colsample\_bytree=0.8

2단계 학습: 트리 최적화 → sample\_weight 적용

- LGBMRegressor (27개 피쳐)

파라미터: learning\_rate=0.05, max\_depth=5, num\_leaves=50, subsample=0.9

2단계 학습: 트리 최적화 → 가중치 적용

- CatBoostRegressor (32개 피쳐)

파라미터: learning\_rate=0.1, depth=5, subsample=0.9, colsample\_bylevel=0.8

2단계 학습: 트리 최적화 → 가중치 적용

## 최종 모델(메타 모델)

- Ridge Regression














Base Model들의 예측을 선형 결합하면서 정규화를 통해 안정성을 확보하기 위함

# 모델\_Ensemble

## Model\_Ensemble

리더보드 기준으로 앙상블 후보를 선정 후 가중평균 앙상블을 통해 최종 제출물을 생성

앙상블 목록	LB Score
Catboost 단일 모델	5.29539
lightgbm 단일 모델	5.39090
xgboost 단일 모델	5.39160
catboost 유형별 모델	5.78100
lightgbm 유형별 모델	5.81680
xgboost 유형별 모델	5.7983
comparison	5.56768
comparison2	5.22
Catboost 알파값 앙상블	5.3765
Staking	5.2633

14	필_	  	5.75511	76
15	민규정		5.76176	18
16	블랙		5.7712	3
17	혈		5.77963	6
18	마약근절	 	5.82404	127
19	전기적 참견 시점	  	5.8266	46
20	tghwang		5.83877	67
21	rlawhdqls		5.86276	61



감사합니다.

서동현, 유동현

