

지역난방 열수요 예측 보고서

참가번호 250275 팀명 대충읽음

1. 문제 정의 및 연구 목적

지역난방 시스템은 도시 기반 에너지 인프라의 핵심으로, 주거 및 상업시설에 열에너지를 공급하는 데 중요한 역할을 한다. 열수요는 외기온, 풍속, 시간대, 계절 등 다양한 기상 및 시간 요인에 따라 비선형적으로 변화하며, 실시간 수요의 변동성이 큰 특징을 지닌다. 이에 따라 안정적인 열공급과 설비 운영 효율화를 위해서는 정밀한 수요 예측이 필수적이다.

본 연구의 목적은 기상 관측 데이터와 시간 정보를 기반으로, 공급 지점 단위의 시간별 열수요를 예측하는 기계학습 기반 모델을 구축하는 것이다. 이를 위해 트리 기반 회귀 모델들을 결합한 앙상블 모델(Stacking Regressor)을 적용하였으며, 난방 시즌 여부에 따라 데이터를 난방기/비난방기로 분할하여 각각의 모델을 학습함으로써 계절적 특성을 효과적으로 반영하였다.

예측 성능 향상을 위해 체감온도, 난방도일, 시차 변수, 이동평균, 공휴일 여부, 이상기상 플래그 등 도메인 지식에 기반한 다양한 파생변수를 설계하였다. 모델의 하이퍼파라미터는 Optuna 프레임워크를 활용하여 자동으로 최적화함으로써, 과적합을 방지하고 일반화 성능을 극대화하였다.

제안한 예측 모델은 에너지 수급 계획 수립, 열공급 설비 운영 최적화, 피크 수요 대응 등 다양한 실무 영역에서의 효율성 향상에 기여할 수 있다. 또한, 기후 변화나 에너지 수급의 불확실성에 유연하게 대응할 수 있는 기반 기술로서 활용 가능성이 있다.

2. 분석 데이터 및 전처리

2.1 주요 변수 소개

본 연구에 사용된 데이터는 시간 단위로 수집된 지역난방 열수요량(heat_demand)과 기상 데이터를 포함하고 있으며, 총 18개 열공급 지점(branch_id)에 대한 관측값으로 구성되어 있다. 주요 기상 변수로는 기온(ta), 습도(hm), 풍속(ws), 풍향(wd), 강수량(rn_day) 등이 있으며, 각 시간대별 수요 예측을 위한 충분한 시계열적, 기상적 정보를 포함한다.

변수명	정의	변수명	정의
tm	시간(YYYY-MM-DD-HH)	dayofweek	요일 (0:월~6:일)
branch_ID	지사명 식별자	dayofyear	연중 일수
ta	기온(°C)	hour_sin	시간의 sin 변환
wd	풍향(도)	hour_cos	시간의 cos 변환
ws	풍속(m/s)	month_sin	월의 sin 변환
rn_day	일 강수량 (mm)	month_cos	월의 cos 변환
rn_hr1	시간 강수량 (mm/h)	dayofweek_sin	요일의 sin 변환
hm	상대습도 (%)	dayofweek_cos	요일의 cos 변환
si	일사량 (MJ/m^2)	HDD18	난방도일 (18°C 기준)
ta_chi	체감온도	apparent_temp	고급 체감온도
ta_missing	기온 결측치 여부	ta_lag_3h	기온 3시간 시차
ws_missing	풍속 결측치 여부	ta_lag_6h	기온 6시간 시차
rn_day_missing	일 강수량 결측치 여부	ta_lag_24h	기온 24시간 시차
rn_hr1_missing	시간 강수량 결측치 여부	ta_ma_6h	기온 6시간 이동 평균
hm_missing	습도 결측치 여부	ta_ma_12h	기온 12시간 이동 평균
si_missing	일사량 결측치 여부	ta_ma_24h	기온 24시간 이동 평균
ta_chi_missing	체감온도 결측치 여부	ta_diff_3h	기온 3시간 차이

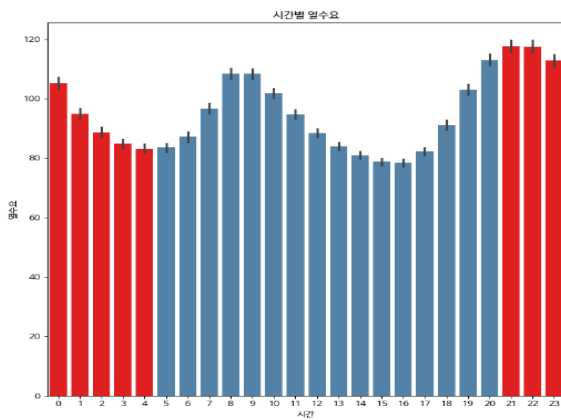
heat_demand_missing	열수요 결측치 여부	ta_diff_6h	기온 6시간 차이
month	월 (1~12)	is_holiday	공휴일 여부
day	일	heating_season	난방 시즌 여부 (10~4월)
hour	시간 (0~23)	cold_extreme	기온 하위 10%
strong_wind	풍속 상위 10%	heavy_rain	강수 상위 10%

[표 1] 사용된 변수

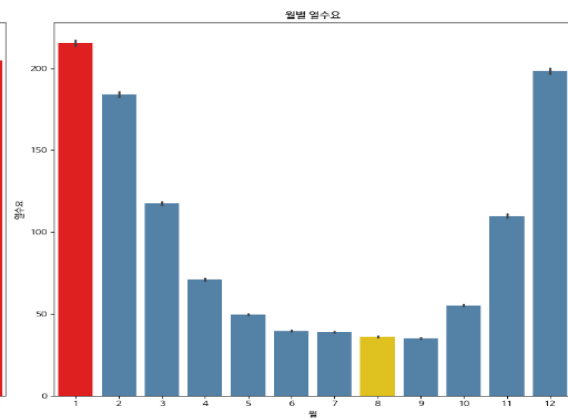
2.2 시계열 및 계절성 분석

열수요 데이터는 시간 흐름에 따른 시계열적 특성과 뚜렷한 계절성을 동시에 지닌다. 이를 효과적으로 모델링하기 위해 시간 파생 변수와 주기 함수 변수를 활용한 전처리를 수행하였다.

먼저, 시간 단위별 열수요의 평균을 분석한 결과, 하루 중 오전(6~9시)과 저녁(18~21시)에 수요가 집중되는 양상을 보였다. 이는 일상적인 난방 사용 시간과 일치하는 경향으로 해석할 수 있다. 월별 평균 열수요를 확인한 결과, 1월과 12월 등 겨울철에 수요가 크게 증가하고, 여름철(7~8월)에는 감소하는 뚜렷한 계절성 패턴이 관찰되었다.



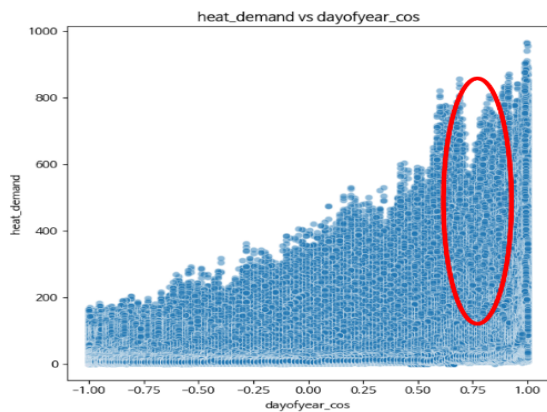
[그림 1]: 시간대별 열수요 평균 (Barplot)



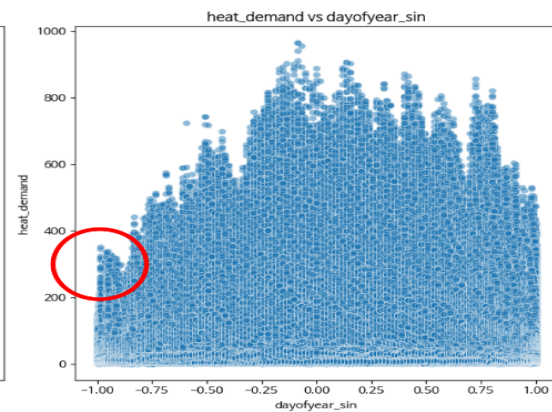
[그림 2]: 월별 열수요 평균 (Barplot)

이러한 반복적인 시간 및 계절 흐름을 모델에 자연스럽게 반영하기 위해, 사인(sin)과 코사인(cos) 기반의 주기 함수 파생 변수를 도입하였다. 연간 주기를 표현하기 위해, day_of_year 변수를 기준으로 1년 주기의 sin/cos 값을 계산하여 연간 계절성을 수치화하였다. 이 방법은 월(month) 또는 요일(weekday) 변수보다 더 부드럽게 순환 패턴을 포착할 수 있다. 동일한 방식으로 hour 변수에도 sin/cos 변환을 적용하여, 하루 단위 주기성을 반영하였다.

또한, 열수요는 계절에 따라 뚜렷한 수요 차이를 보이므로, 난방 수요가 급증하는 10월~4월을 난방 시즌으로 정의하고 해당 여부를 나타내는 heating_season 플래그 변수를 생성하였다. 이후, 난방기와 비난방기의 열수요 패턴을 고려하여 전체 데이터를 두 그룹으로 나누고, 각 시즌별로 모델을 분리 학습하는 전략을 채택하였다. 이로써 시간 데이터를 보다 효과적으로 학습하고 예측 성능을 향상시킬 수 있었다.



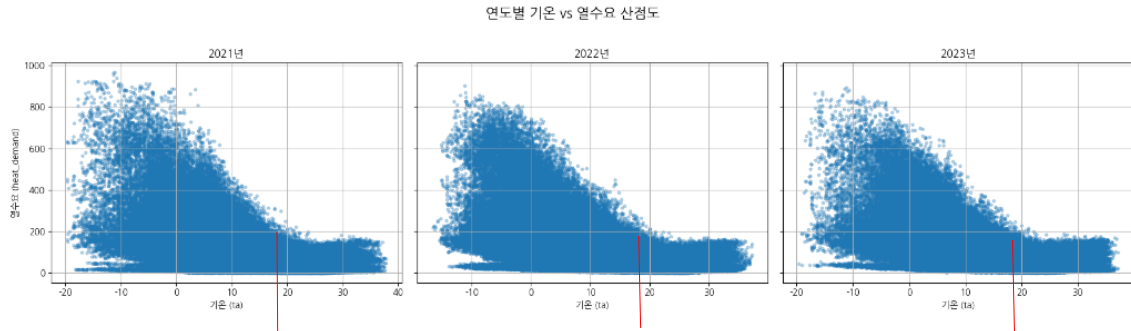
[그림 3]: dayofyear_cos (산점도)



[그림 4]: dayofyear_sin (산점도)

cos 값이 1에 가까운 겨울철에 열수요가 뚜렷하게 증가하고, sin 값이 큰 여름철에는 수요가 낮아지는 비선형 패턴이 뚜렷하게 관측되었다. 이는 단순 시간 변수보다 주기성을 반영한 함수가 예측 모델에 유효함을 시사한다.

2.3 임계 기온(18°C) 기반 열 수요 분석



[그림 5]: 연도별 & 기온별 열수요 산점도

기온과 열수요 간의 관계를 분석한 결과, 외기온이 약 18°C를 기준으로 열수요 변화에 명확한 분기점을 형성함을 확인하였다. 특히 18°C 이하에서는 기온이 낮아질수록 열수요가 급격히 증가하는 경향을 보였다. 이는 지역난방의 난방 개시 기준과도 부합하는 값으로, 기온 18°C를 임계값으로 설정한 HDD18(Heating Degree Day) 지표를 파생 변수로 도입하였다.

HDD18은 '18°C - 기온' 값을 기반으로 하며, 기온이 18°C 미만일 때만 양의 값을 갖는다. 해당 변수는 저온 구간에서의 열수요 민감도를 모델이 더 정밀하게 학습할 수 있도록 도와주며, 특히 난방 시즌 중 급격한 수요 상승 구간을 효과적으로 반영할 수 있는 수단이 된다.

2.4 시차 및 이동 평균 변수 활용

열수요는 현재 기온뿐만 아니라 직전 시간대의 기온 흐름에도 민감하게 반응하는 경향이 있다. 이에 따라 시차(lag), 이동 평균(moving average), 변화량(diff) 시계열 기반 파생 변수를 도입하여 열 수요의 시간적 연속성과 동적 변화를 모델에 반영하고자 하였다.

기온의 시간 지연 효과를 반영하기 위해, 3시간, 6시간, 24시간 전의 기온을 나타내는 ta_lag_3h, ta_lag_6h, ta_lag_24h 변수를 생성하여 시간 지연 효과를 고려하였다. 또한, 기온의 흐름을 부드럽게 반영하기 위해 일정 시간 구간의 평균 기온을 나타내는 ta_ma_6h, ta_ma_12h, ta_ma_24h 변수를 도입하였다. 이와 함께, 단시간 내 기온 변화 폭이 수요에 미치는 영향을 반영하기 위해, 최근 3시간 및 6시간 동안의 기온 차이를 나타내는 ta_diff_3h, ta_diff_6h 변수도 추가하였다.

이러한 변수들은 모두 지사별(branch_id 기준)로 계산되었으며, 열수요의 시간적 연속성과 변화에 대한 단기 추세 변화를 효과적으로 포착하는데 기여하였다.

2.5 체감온도 반영을 통한 열수요 영향 분석

열수요는 단순한 기온 수치보다, 사람이 체감하는 온도에 더 민감하게 반응한다. 특히 동일한 기온이라도 풍속이 강할 경우 더 춥게 느껴지며, 이는 난방 수요 증가로 이어질 수 있다. 이러한 체감 추위의 효과를 반영하기 위해, 본 연구에서는 기온(ta)과 풍속(ws)을 바탕으로 캐나다 기상청 공식을 활용한 체감온도(apparent temperature) 변수를 새롭게 도입하였다.

$$13.12 + 0.6215 \times ta - 11.37 \times (ws \times 3.6)^{0.16} + 0.3965 \times ta \times (ws \times 3.6)^{0.16}$$

이 변수는 특히 겨울철 강풍 상황에서 실제보다 더 크게 증가하는 난방 수요를 보다 정밀하게 반영하는 데 유효하며, 기존 기온 변수의 한계를 보완하는 수단으로 활용하였다.

2.6 열수요에 영향을 주는 비정상 요인 분석

공휴일, 한파, 강풍, 호우와 같은 비정상적인 외부 조건은 일반적인 기상·시간 요인과는 다른 방

식으로 열수요에 영향을 줄 수 있다. 이에 따라, 각 조건이 열수요에 미치는 영향을 정량적으로 파악하기 위해, 해당 조건별로 평균, 분산, 표준편차를 산출하고 그룹 간 차이에 대한 통계적 유의성 검정을 수행하였다.

이상치 조건별 데이터는 정상일에 비해 상대적으로 표본 수가 적고 등분산성이 확보되지 않을 가능성이 존재하기 때문에, 평균 비교에는 Welch's t-test를 사용하여 보다 신뢰성 있는 통계 결과를 확보하였다.

	구분	열수요 평균	표준편차	분산	p-value
0	평일	95.900000	114.700000	13151.500000	0.577700
1	공휴일	95.400000	125.800000	15829.300000	0.577700
2	정상	88.300000	103.000000	10614.000000	0.000000
3	한파	166.700000	181.200000	32848.300000	0.000000
4	정상	95.300000	114.700000	13160.300000	0.000000
5	강풍	101.300000	120.400000	14494.200000	0.000000
6	정상	97.000000	116.900000	13654.800000	0.000000
7	호우	84.400000	97.200000	9449.300000	0.000000

[그림 6]: 조건별 열수요 평균 차이와 통계적 유의성

공휴일, 한파, 강풍, 호우와 같은 비정상적인 외부 조건은 일반적인 기상·시간 요인과는 다른 방식으로 열수요에 영향을 줄 수 있다. 이에 따라, 각 조건이 열수요에 미치는 영향을 정량적으로 파악하기 위해, 해당 조건별로 평균, 분산, 표준편차를 산출하고 그룹 간 차이에 대한 통계적 유의성 검정을 수행하였다.

이상치 조건별 데이터는 정상일에 비해 상대적으로 표본 수가 적고 등분산성이 확보되지 않을 가능성이 존재하기 때문에, 평균 비교에는 Welch's t-test를 사용하여 보다 신뢰성 있는 통계 결과를 확보하였다.

이에 따라 is_holiday, cold_extreme, strong_wind, heavy_rain과 같은 기상 플래그 변수를 생성하였다. 이는 모델이 통상적인 예측 조건을 넘어서는 비정상 상황에서의 열수요 변화를 더 유연하게 학습할 수 있도록 한다. 결과적으로 예측 정확도 향상 및 안정적 운영 의사결정에 기여할 수 있을 것으로 기대된다.

3. 모델선정

열수요 예측 문제는 시간 흐름에 따른 비선형적 변화뿐 아니라, 기상 요인과 공간 정보 간의 복잡한 상호작용을 수반한다. 이러한 특성은 단순한 선형 회귀 기반 모델만으로는 충분히 포착하기 어려우며, 다양한 변수 간의 복합적 영향력을 반영할 수 있는 기법이 필요하다.

이에 따라 본 연구에서는 LightGBM, XGBoost 등 트리 기반 기법을 결합한 앙상블 회귀 모델을 구성하였다. 이 모델은 개별 기법의 예측값을 종합하는 스택킹 방식으로 구성되어 있으며, 비선형성, 변수 간 상호작용, 이상치 및 결측치에 대한 견고한 처리가 가능하다는 장점을 갖는다.

또한 열수요는 난방 시즌(10월~4월)과 비난방 시즌(5월~9월) 간 수요 구조가 뚜렷하게 다르므로, 두 시기를 별도 그룹으로 구분하여 계절별 특화 모델을 학습하였다. 이 과정에서 각 모델의 하이퍼파라미터는 Optuna 프레임워크를 활용해 자동으로 최적화함으로써, 과적합을 방지하고 일반화 성능을 극대화하였다.

이러한 설계는 계절성과 기상 특성의 영향을 보다 정밀하게 반영하고, 예측의 안정성과 정확도를 동시에 향상시키는 데 기여하였다.

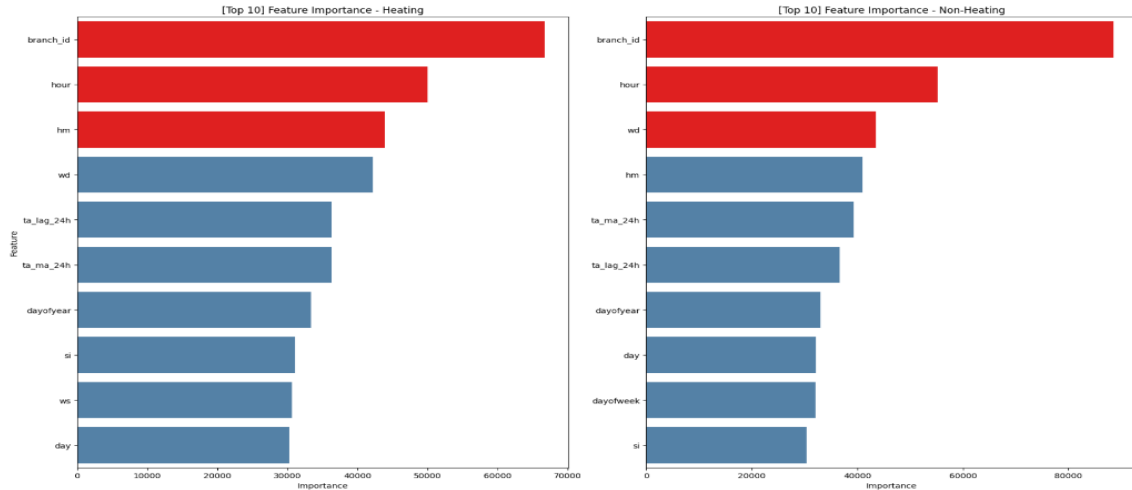
4. 결과 및 검증

4.1 성능평가

본 예측 모델은 시계열성과 기상·공간 정보를 복합적으로 반영한 구조로서, 비선형성과 변수 간 상호작용을 효과적으로 포착할 수 있는 트리 기반 모델인 LightGBM을 기반으로 개발하였다.

모델 성능 평가는 다음 세 가지 지표를 기준으로 수행되었다:

- RMSE (Root Mean Squared Error)
- MAE (Mean Absolute Error)
- R^2 (Coefficient of Determination)



[그림 7]: 변수 중요도

변수 중요도 분석 결과, 공간적 식별자인 branch_id, 시간 변수(hour, month), 습도(hm), 기온(ta), 체감온도(ta_chi) 등이 상위권에 포함되었다. 이는 열수요가 지점의 위치적 특성, 시간대, 기온/습도 조합에 따라 크게 달라진다는 점을 시사하며, 모델이 주요 변수의 기여도를 효과적으로 학습하고 있음을 보여준다.

4.2 예측 결과 해석

최종 예측 모델은 트리 기반의 LightGBM 회귀 모델을 사용하였다. LightGBM은 비선형적 관계 및 변수 간의 복잡한 상호작용을 효과적으로 학습할 수 있는 장점이 있어, 본 연구의 열수요 예측 문제에 적합한 모델로 선정되었다. 또한 난방기와 비난방기의 계절적 수요 차이를 반영하기 위해, 데이터를 난방 시즌과 비난방 시즌으로 나누어 각각 별도의 모델을 학습하였다.

구분	RMSE	MAE	R ²
HEATING	11.1096	5.2924	0.9611
NON-HEATING	6.3131	3.5477	0.9933

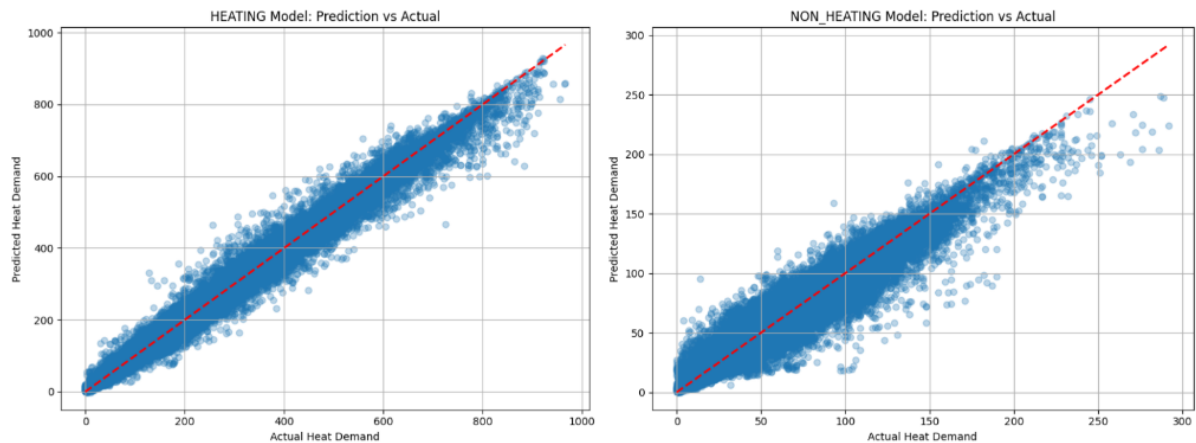
[표 2] 그룹 별 예측 결과

예측 결과는 대각선을 중심으로 실제값과의 정합성이 높았으며, 전반적으로 이상치 없이 안정적인 성능을 나타냈다. 특히 비난방시즌의 정확도는 0.99로, 예측값이 실제 관측값을 정밀하게 설명하고 있음을 시사한다.

또한 시간대 및 계절 정보에 대한 파생 변수를 설계함으로써, 특정 시간(예: 출근 전후), 특정 계절(예: 한파기간)의 수요 급등 현상을 효과적으로 포착하였다.

이는 단순 시계열 기반 모델 대비 기상-시간 융합 구조가 예측력 향상에 실질적 기여를 하였음을 의미한다.

한편, 난방 시즌의 RMSE가 상대적으로 더 높은 것은 급격한 기온 하강이나 한파와 같은 이상기상



[그림 8]: 예측값 vs 실제값 산점도

상황에서 수요가 급증하는 경향 때문으로 판단된다.

이는 실제 수요가 순간적으로 큰 폭으로 변동하기 때문에 예측 모델이 완전히 추적하지 못한 결과로 해석할 수 있으며, 이는 본 모델의 한계이자 향후 개선 여지로도 볼 수 있다.

반면, 비난방 시즌의 경우 상대적으로 기온 변화가 완만하고 수요 패턴이 안정적인 특성을 보이므로 예측 정확도(R^2 0.99 이상)가 높게 나타난 것으로 분석된다.

이는 열수요 패턴이 계절적 요인에 따라 명확히 구분되며, 시즌별 모델 학습이 예측 성능 향상에 기여했음을 의미한다.

참가번호 250275 의
RMSE 는 15.73 입니다.

실제 테스트셋을 이용한 검증 결과는

5. 활용방안

본 연구에서는 기상청의 시계열 및 공간 데이터를 기반으로, 지역난방 열수요를 정밀하게 예측하는 회귀 모델을 구축하였다.

시간 정보로부터 유도된 주기성 파생 변수(hour_sin, hour_cos, month_sin, month_cos, dayofweek_sin, dayofweek_cos), 기상 변수(ta, ws, rn_day, hm 등)의 결측 여부, 그리고 시차 및 이동 평균 기반의 파생 변수(ta_lag_3h, ta_ma_6h, ta_diff_3h 등), 이벤트 및 이상치 플래그 변수(is_holiday, cold_extreme, strong_wind, heavy_rain)를 종합적으로 활용하여 입력 특성을 설계하였다.

예측 모델은 LightGBM을 기반으로 한 앙상블 회귀 구조로 구성되었으며, 난방 시즌(heating)과 비난방 시즌(non-heating)을 구분하여 그룹별로 개별 모델을 학습하였다. 각 모델은 Optuna 기반의 하이퍼파라미터 튜닝을 통해 최적화되었고, 최종 예측은 여러 모델의 결과를 평균하는 방식으로 통합함으로써 과적합을 방지하고 예측 안정성을 확보하였다.

모델은 주요 평가 지표(RMSE, MAE, R^2)에서 우수한 예측력과 일반화 성능을 보였으며, 풍향, 기온, 강수량 등 주요 변수들의 영향을 정량적으로 분석함으로써 열수요 변동의 원인 파악에도 기여하였다.

제안된 모델은 향후 실시간 수요 반영형 난방 운영 시스템, 기상 예보 기반의 선제적 열공급 계획 수립 등 다양한 실무 현장에 직접 활용될 수 있으며, 기상 예측 시스템과의 융합을 통한 확장 가능성 또한 높다.