지역난방 열수요와 날씨 빅데이터를 융합한 열수요 예측

원자력발전소

250340



분석 배경

- 열수요란?

가정과 상업공간 등에서 난방을 위해 필요한 열 공급량을 의미함 일반적으로 외기온도, 사용자의 활동, 패턴, 외부 환경에 많은 영향을 받음

- 배경



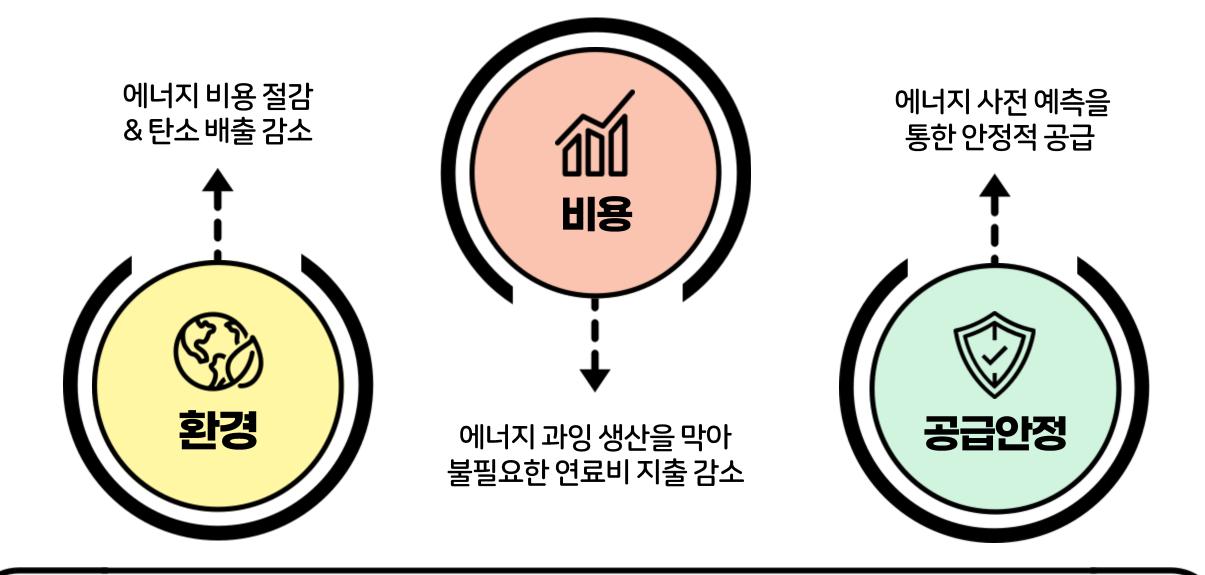


열수요 예측 모형을 통해 잠재적으로 발생할 수 있는 문제 해결과 안정적인 열 공급, 요금 안정이 기대됨



분석배경및목표

- 목표



- 1. 주어진 기상, 열수요 데이터를 활용하여 각 지사들의 열수요를 예측하는 모형 개발
- 2. 열수요 예측을 이용한 활용방안 제시



분석 데이터

- 변수

변수명	설명	변수명	설명
tm	날짜 및 시간	rn_hr1	1시간 강수량
branch_ID	지사명	hm	정시 상대 습도
ta	정시 기온	si	ASOS 일사량
wd	정시 10분 평균 풍향	ta_chi	체감온도
WS	정시 10분 평균 풍속	heat_demand	시간당 지사별 열 공급량
rn_day	해당 시간까지의 일 강수량		

- 1. rn_hr1의 정보안에 rn_day의 정보가 중복되어 있어, 다중 공선성으로 인해 rn_day는 사용하지 않음
- 2. 변수 중 ta, ws, rn_hr1, hm, si, ta_chi에 결측치가 존재했으며, -9.9, -99.9로 표현됨. 이를 처리하기 위해 NA로 표현한 뒤 분석을 진행



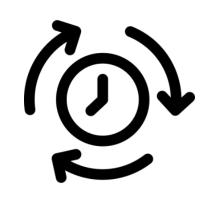
데이터 전처리 - 파생변수

- 시간적 파생변수



주말/ 휴일

: 주말과 법정 공휴일 여부를 나타냄



시간 주기성

: 달, 월, 일에 대해 삼각 함수 변환을 통해 시간의 주기성을 표현



지연 변수

: 기온, 상대습도, 풍속 변수의 과거 값을 현재 시점의 설명변수로 활용하는

지연 변수를 생성하여 예측에 사용



데이터 전처리 - 파생변수

- 환경적 파생변수



난방도일(Hour, 7D, 30D)

: 특정지역의 난방 필요량을 나타내는 지표 (기준온도 18°C)



불쾌지수

DI = $1.8 \times \text{Ta} - 0.55(1 - \text{RH})(1.8 \times \text{Ta} - 26) + 32$

Ta: 건구온도(°C)

RH: 상대습도

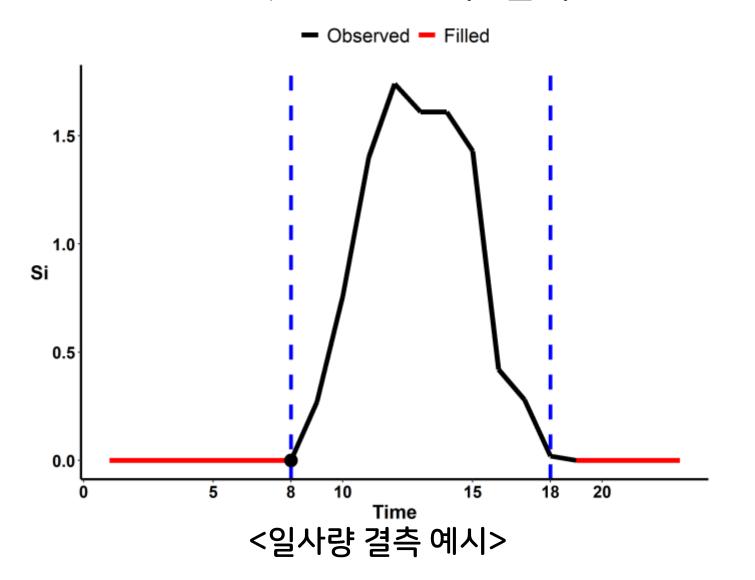


데이터 전처리 - 결측치 보간

- Case1: 비무작위 결측

비무작위 결측(MNAR) 문제 : 일사량

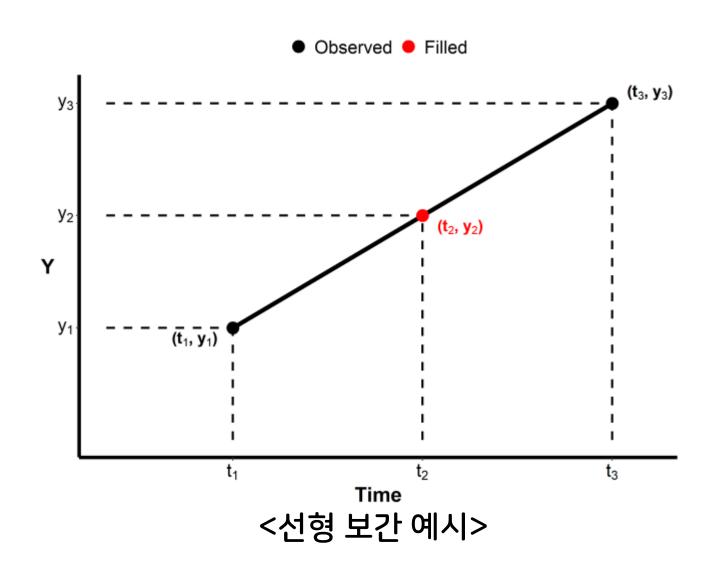
→ 08:00 이전, 18:00 이후 결측치 0으로 대체



- Case2: 단기 결측

1개 또는 2개 연속 결측

→ 전, 후 값을 이용해 선형 보간 진행





데이터 전처리 - 결측치 보간

- Case3: 장기 결측

3개 이상의 연속 결측

- → 결측 직전 시점의 값과 클러스터내 평균 이용한 보간 진행
- → 과거 시점의 정보인 직전 값과 클러스터 내 평균값으로 해당 시점의 정보를 반영

$$y_t = \frac{y_{t-1} + c_t}{2}$$

 y_t : 결측값

 y_{t-1} : 결측 직전 시점의 값

 c_t : 클러스터 내 평균값

$$y_{t-1} \longrightarrow \widehat{y}_t \longrightarrow \widehat{y}_{t+1} \longrightarrow \widehat{y}_{t+2}$$

$$\uparrow \qquad \uparrow \qquad \uparrow \qquad \uparrow \qquad c_{t+1} \qquad c_{t+2}$$

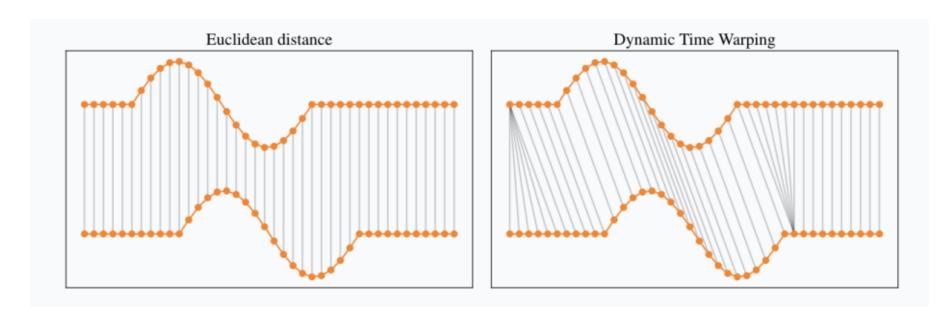
<결측치 보간 예시>



데이터 전처리 - 클러스터링

- 지사별 거리 기반 유사도

DTW(Dynamic Time Wraping)를 통한 시계열 유사도 분석 후 지사 클러스터링



<Euclidean vs DTW>

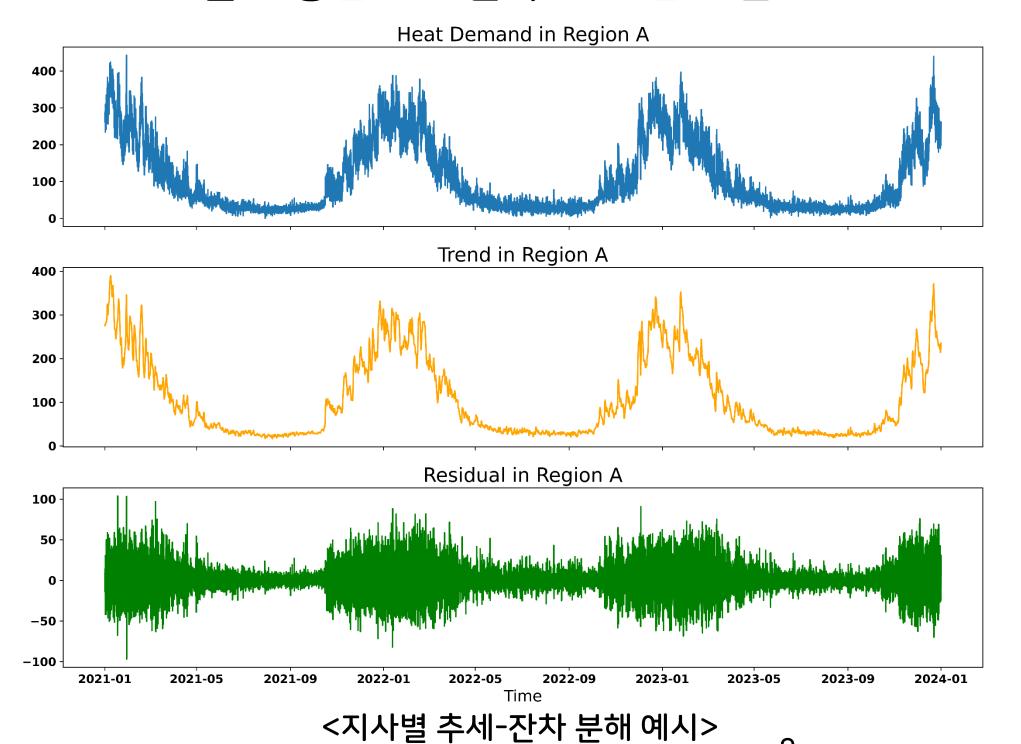
- DTW 기반 계층적 클러스터링





데이터 전처리 - 분해

- HP Filter를 이용한 지사별 추세와 잔차 분해



HP Filter:

$$\min_{\tau_t} \sum_{t=1}^{T} (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} (\Delta^2 \tau_t)^2$$

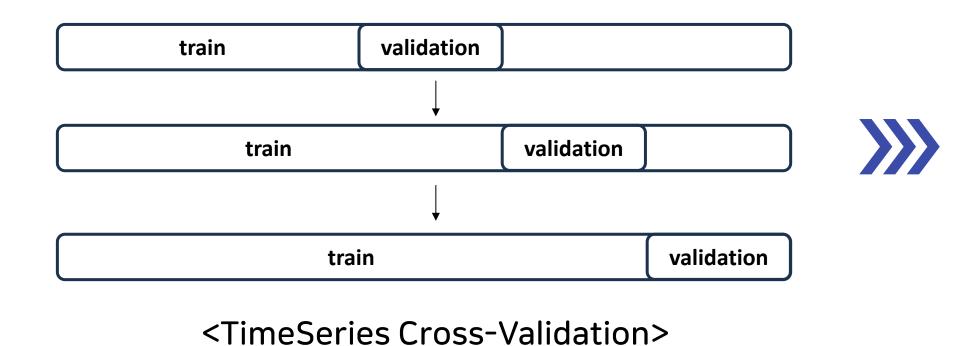
 τ_t : 추세

$$r_t = y_t - \tau_t$$
: 잔차



모델별최적화

- 하이퍼파라미터 선택



파라미터	범위(분포)	
n_estimators	500 - 5000 (step 100)	
max_depth	3 - 12 (정수)	
learning_rate	0.01 - 0.30 (log-uniform)	
subsample, colsample_bytree, colsample_bylevel	0.70 - 1.00 (연속)	
reg_alpha, reg_lambda	1e-8 – 1.0 (log-uniform)	
min_child_weight	1 – 7	
gamma	1e-8 – 0.5 (log-uniform)	

<하이퍼파라미터 예시>

파이썬의 optuna를 통해 지사별 RMSE를 최소화하는 방향으로 하이퍼파라미터 탐색 및 모델 최적화 진행



지사별 추세 - 잔차 이중 모델

- 지사별 예측모델

: 추세와 잔차를 지사별로 분리하여 학습

$$\widehat{y_t} = \widehat{\tau_t} + \widehat{r_t}$$

 $\widehat{ au_t}$: 지사별 추세 예측값

 $\widehat{r_t}$: 지사별 잔차 예측값

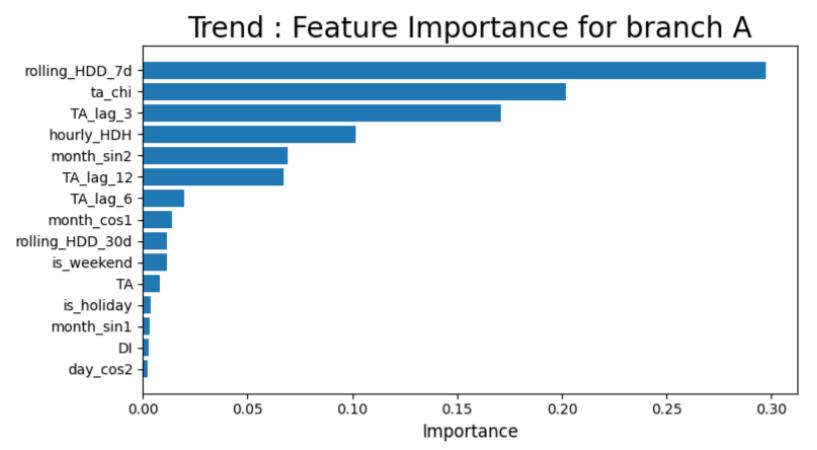
- 모델 선택

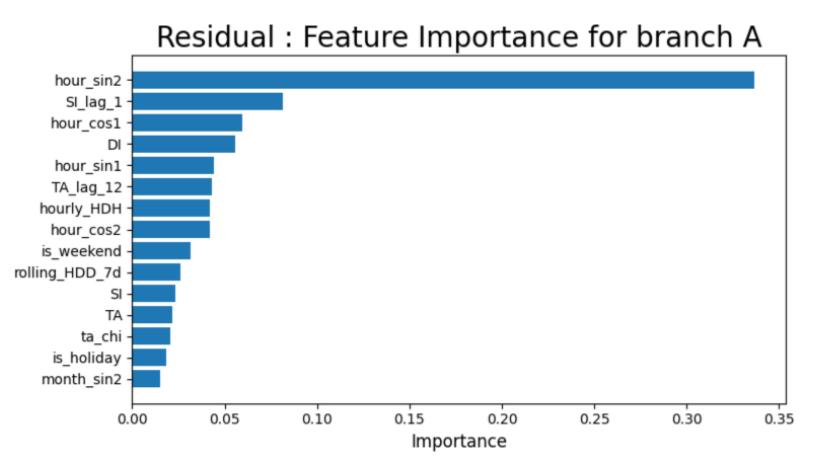
Trend 모델	Residual 모델	RMSE
XGBoost	XGBoost	17.53
XGBoost	Random Forest	19.23
Random Forest	XGBoost	20.84
Random Forest	Random Forest	21.67

2023년을 Validation Set으로 사용해, RMSE가 가장 낮은 XGBoost + XGBoost 모델 선택



결과및해석





<추세 모델 변수 중요도 예시>

<잔차 모델 변수 중요도 예시>

최종 RMSE : 15.53 추세 모델에선 난방도일, 체감 온도, 지연 온도 등이 중요한 변수로

자 모델에선 신경포질, 제임 근모, 지원 근모 등이 공표한 친구모 잔차 모델에선 시간 주기, 지연 일사량, 불쾌 지수 등이 중요한 변수로 파악





에코 마일리지 개선



열수요 예측 모델을 활용해 올해의 날씨를 반영한 동적인 절약 목표 (에코 마일리지, 탄소 포인트 제도 등)의 가이드라인 제시 이를 통해 자발적인 에너지 절약 행동을 유도하고 탄소 배출량 감소에 기여

친환경 계획 수립



정확한 열 수요 예측은 기존 공급 시스템의 효율을 극대화하는 것과 동시에, 미래 에너지 시스템의 핵심인 신재생에너지의 안정적 도입을 위한 '데이터 기반 나침반' 역할을 수행



참고문헌

- 1. 채덕종, [집중분석] 포화수요 도달...변곡점 맞은 지역난방, 이투뉴스 , https://www.e2news.com/news/articleView.html?idxno=301799
- 2. 변국영, "2028년까지 총 446만 세대에 지역난방 공급된다", 에너지데일리 , https://www.energydaily.co.kr/news/articleView.html?idxno=153411
- 3. 서병선, & 심상렬. (2012). 지역난방 열에너지 수요예측. 에너지경제연구, 11(2), 27-55.
- 4. Ravn, Morten O.; Uhlig, Harald (2001): On Adjusting the HP-Filter for the Frequency of Observations, CESifo Working Paper, No. 479, Center for Economic Studies and ifo Institute (CESifo), Munich
- 5. Romain Tavenard (2021), An introduction to Dynamic Time Warping, https://rtavenar.github.io/blog/dtw.html