

AI 기반 건물 에너지 피크 예측 및 다국어 진단 보고서 서비스

정가은 · 전문기 · 맥슈웰 데이브 · 김경현 | 부산외국어대학교 AI 부트 캠프

연구 배경 및 목적

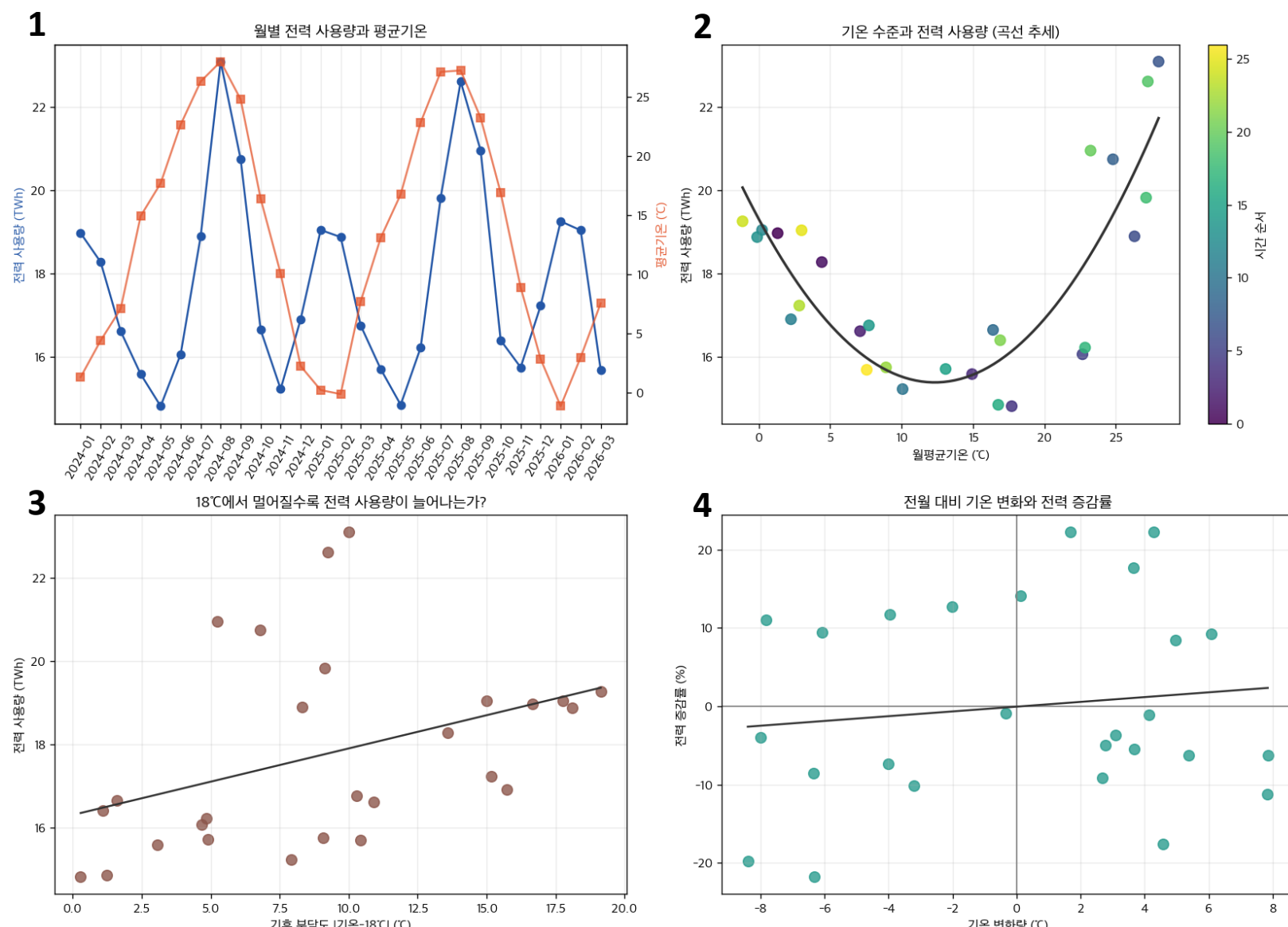
건물주는 월별 사용량의 '피크'와 기준 대비 수준을 한눈에 파악하기 어렵습니다. 공공데이터를 연결해 사용 패턴을 설명하고, 예측 결과와 국내외 기준을 근거로 실행 가능한 다국어 에너지 진단 보고서를 제공하는 것이 목표입니다.

공공데이터 수집 → 분석·시각화 → 머신러닝 → RAG → 다국어 보고서

데이터 수집 및 분석·시각화

건물면적 545,636동 · 건물 전력 40,509,467건
가구 평균 전력 16,304개 · ASOS 95개 지점 89,296일
*ASOS: 기상관측시스템

2021~2026 공공데이터를 정제하고, 전력 사용량이 없는 값(결측치)을 임의로 '0'으로 채우지 않고 그대로 두어, 실제 사용량이 적어 보이는 왜곡 방지

23.10 TWh전체 건물 월별
전력사용량 최고치**상관계수 0.513**18°C 기준 기후부하가 클수록
전력사용량도 늘어나는 경향**70.2 kWh**건물 1㎡당 연간
전력사용량의 중앙값

- 월별 전력사용량과 평균기온의 추이 - 계절에 따라 함께 오르내리는 패턴
- 기온 수준별 전력사용량 - 기온이 18°C에서 멀어질수록(춥거나 더울수록) 사용량 증가(U자형)
- 기후부하(기온-18°C)와 전력사용량의 관계 - 기후부하가 클수록 사용량도 대체로 증가
- 전월 대비 기온 변화량과 전력 증감률 - 뚜렷한 관계 없음(중요한 건 변화량이 아니라 현재 기온 수준)

머신러닝 · 두 예측 단위를 분리해 검증

TRACK A

개별 건물

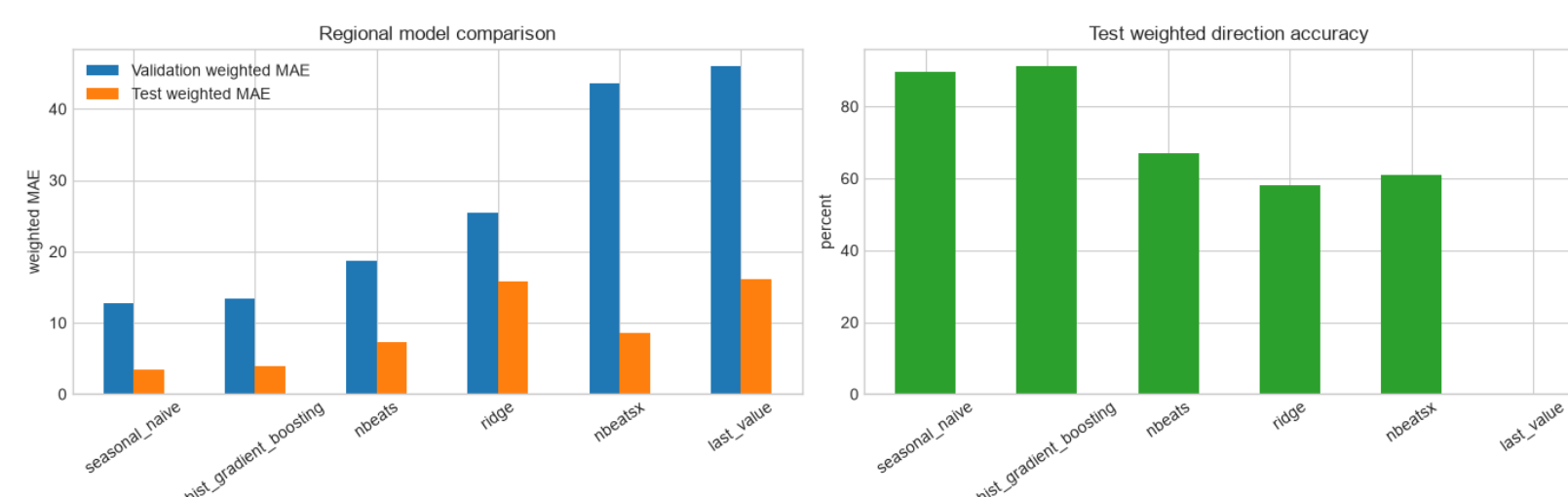
면적·용도·월·기상만으로
월 전력량 예측 · N-BEATSx
31개 변수 · 16만 파라미터

TRACK B

지역 가구 평균

29개월로 6개 모델을 비교
과거값·계절·기상 지연변수 활용
미래 실측 기상은 사용하지 않음

검증지표로 고른 모델은 가장 단순한 계절 기준모델

**6개**

후보 모델 비교

12.78 kWh

검증 가중 MAE

3.45 kWh5개월 테스트 가중
MAE

N-BEATSx 결과

개발셋 RMSE 1.78로 로그 오차는 개선했으나, 예측 총량은 실제의 약 38%에 그침.

정확도 한계로 실제 의사결정에는 미사용. 과거 기록 없는 신축 건물의 참고용 추정치로만 활용.

해석 시 유의사항

- Track A(개별 건물)와 Track B(지역 평균)는 예측 단위가 달라 직접 비교·합산하지 않음
- 테스트 기간은 5개월(2026년 여름철 미포함)로, 장기 성능을 일반화한 결과는 아님

핵심 인사이트 데이터가 짧을수록 강한 기준모델이 딥러닝의 출발점이 된다.

LLM 활용 · 근거를 찾고, 사용자의 언어로 설명

RAG · 검색 근거를 답변에 연결

PDF → OCR → 700토큰 → 하이브리드 검색 → 근거 6개 → 페이지 인용

12개 자동 테스트 통과 · 3쪽 PDF를 3개 청크로 색인
'ZEB 3 최대 완화 비율' 질문에 답변을 13%와 출처 페이지를 함께 제시

다국어 처리 · 번역은 RAG의 경계에서

외국어 자료 → 한국어 색인 | 검색·답변 → 사용자 언어 보고서

한국어·영어·스페인어·프랑스어·중국어·일본어 6개 언어 설계
숫자·단위·ZEB·kWh·kgCO₂·인용표시는 번역하지 않고 그대로 보존

결론 및 기대효과

계절 피크와 에너지 원단위를 빠르게 선별하고, 모델의 불확실성과 규정 근거를 함께 제시해
비용·탄소 절감 의사결정을 지원