

기상 정보에 따른 가정의 에너지 소비량 예측

2020204066 송승규 (팀장)
2024404060 강민혁
2020204021 문하늘
2022204500 무하마드 이르판 나즈미
2022204046 이은강
2021323033 이주희

목차

- 주제 소개 및 배경
 - 데이터 수집 및 EDA
 - 데이터 전처리
 - 모델 학습 및 평가
 - 결과 해석
 - 결론
-

- 기상 정보를 반영한 전력 수요 예측이 필요함

취재K

제주엔 전기가 넘친다? 에너지 전환의 복병 ‘출력제어’

입력 2021.12.08 (08:00)

국제 : 국제일반

'암흑세계' 겪은 뉴욕...'대정전' 지속되면 무슨 일이 벌어질까

중앙일보 | 입력 2019.07.15 18:35 업데이트 2019.07.15 19:57



일별, 월별 수요량을
정확히 예측하는 것이 중요

- 기상 정보를 반영한 전력 수요 예측이 필요함

냉방도일과 난방도일만 사용



20%의 에너지 낭비



기상 정보 기반의 전력 수요 예측 모델 개발 필요

“기상정보에 따른 가정의 에너지 소비량 예측”

스마트홈 설계에 활용할 가능성 모색

에너지 효율성을 높이는
맞춤형 기술 제안 가능

개인화된 에너지 절약 서비스 개발 및
지역별 에너지정책 수립에 기여

데이터 수집

- “Smart Home Dataset with weather Information” (기상 정보가 포함된 스마트 홈 데이터 세트)

kaggle <https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information>



time	use [kW]	gen [kW]	House overall [kW]	Dishwasher [kW]	Furnace 1 [kW]	Furnace 2 [kW]	Home office [kW]	Fridge [kW]	Wine cellar [kW]	...	visibility	summary	apparent
1451624400	0.932833	0.003483	0.932833	0.000033	0.020700	0.061917	0.442633	0.124150	0.006983	...	10.0	Clear	
1451624401	0.934333	0.003467	0.934333	0.000000	0.020717	0.063817	0.444067	0.124000	0.006983	...	10.0	Clear	
1451624402	0.931817	0.003467	0.931817	0.000017	0.020700	0.062317	0.446067	0.123533	0.006983	...	10.0	Clear	
1451624403	1.022050	0.003483	1.022050	0.000017	0.106900	0.068517	0.446583	0.123133	0.006983	...	10.0	Clear	
1451624404	1.139400	0.003467	1.139400	0.000133	0.236933	0.063983	0.446533	0.122850	0.006850	...	10.0	Clear	

503911 Samples, 32 Columns

데이터 수집

- feature 소개 - 스마트홈 관련

타임스탬프	use [kW]	총 생산 전력량 [kW]	House overall [kW]	식기세척기 [kW]	난방시스템1 [kW]	난방시스템2 [kW]	Home office [kW]	냉장고 [kW]	와인 저장고 [kW]
1451624400	0.934333	0.003483	0.934333	0.000033	0.020700	0.061917	0.444067	0.124150	0.006983
1451624401	0.934333	0.003467	0.934333	0.000000	0.020717	0.063817	0.444067	0.124000	0.006983
1451624402	0.931817	0.003467	0.931817	0.000017	0.020700	0.062317	0.446067	0.123533	0.006983

데이터 수집

● feature 소개 - 스마트홈 관련

차고 문 개폐		주방 장치 14		헤파		전자레인지		생성된 태양광 전력
Garage door [kW]	Kitchen 12 [kW]	Kitchen 14 [kW]	Kitchen 38 [kW]	Barn [kW]	Well [kW]	Microwave [kW]	Living room [kW]	Solar [kW]
503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000
0.014139	0.002755	0.007023	0.000009	0.058530	0.015642	0.010983	0.035313	0.076229
0.014292	0.021860	0.076740	0.000010	0.202706	0.137841	0.098859	0.096056	0.128428

데이터 수집

- feature 소개 - 기상 관련

온도		가시거리		대기압		풍향	
temperature	humidity	visibility	apparentTemperature	pressure	windSpeed	windBearing	precipIntensity
503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000	503910.000000
50.741935	0.664085	9.253444	48.263382	1016.301625	6.649936	202.356843	0.002598
19.113807	0.194389	1.611186	22.027916	7.895185	3.982716	106.520474	0.011257

● 기본적인 탐색적 데이터 분석

time	503910	non-null	object
use [kW]	503910	non-null	float64
gen [kW]	503910	non-null	float64
House overall [kW]	503910	non-null	float64
Dishwasher [kW]	503910	non-null	float64
Furnace 1 [kW]	503910	non-null	float64
Furnace 2 [kW]	503910	non-null	float64
Home office [kW]	503910	non-null	float64
Fridge [kW]	503910	non-null	float64
Wine cellar [kW]	503910	non-null	float64
Garage door [kW]	503910	non-null	float64
Kitchen 12 [kW]	503910	non-null	float64
Kitchen 14 [kW]	503910	non-null	float64
Kitchen 38 [kW]	503910	non-null	float64
Barn [kW]	503910	non-null	float64
Well [kW]	503910	non-null	float64
Microwave [kW]	503910	non-null	float64
Living room [kW]	503910	non-null	float64
Solar [kW]	503910	non-null	float64
temperature	503910	non-null	float64
icon	503910	non-null	object
humidity	503910	non-null	float64
visibility	503910	non-null	float64
summary	503910	non-null	object
apparentTemperature	503910	non-null	float64
pressure	503910	non-null	float64
windSpeed	503910	non-null	float64
cloudCover	503910	non-null	object
windBearing	503910	non-null	float64
precipIntensity	503910	non-null	float64
dewPoint	503910	non-null	float64
precipProbability	503910	non-null	float64

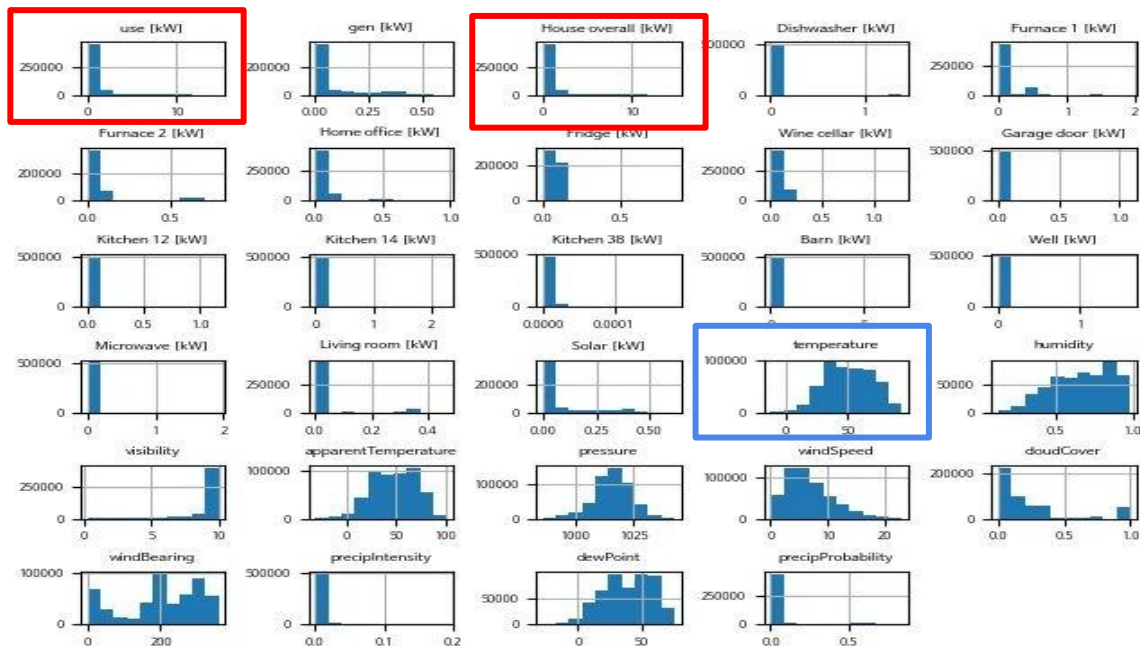
cloudCover 변수

-> 범주형 변수가 아닌 수치형 변수

결측값 존재

-> 제거

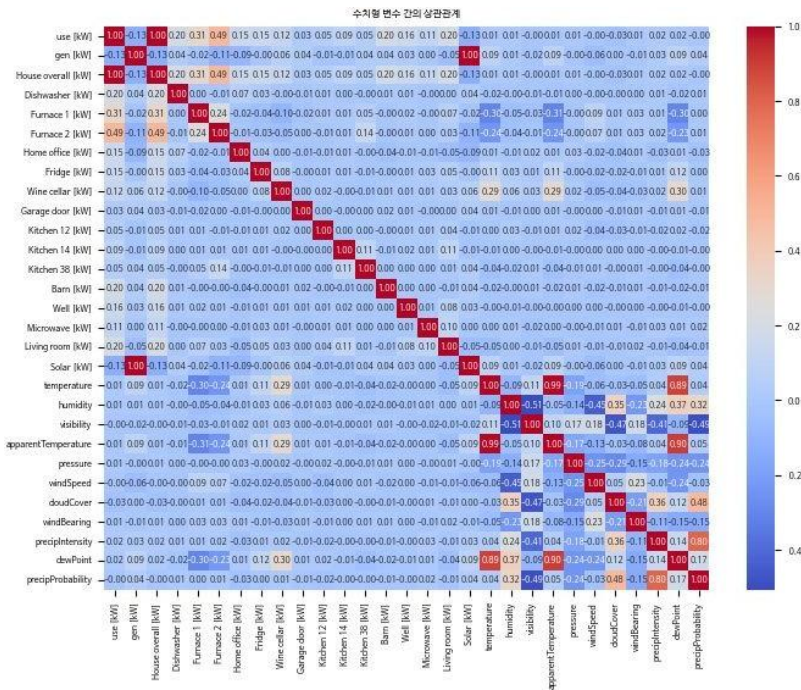
● 변수 분포 및 변수 간 상관관계 확인



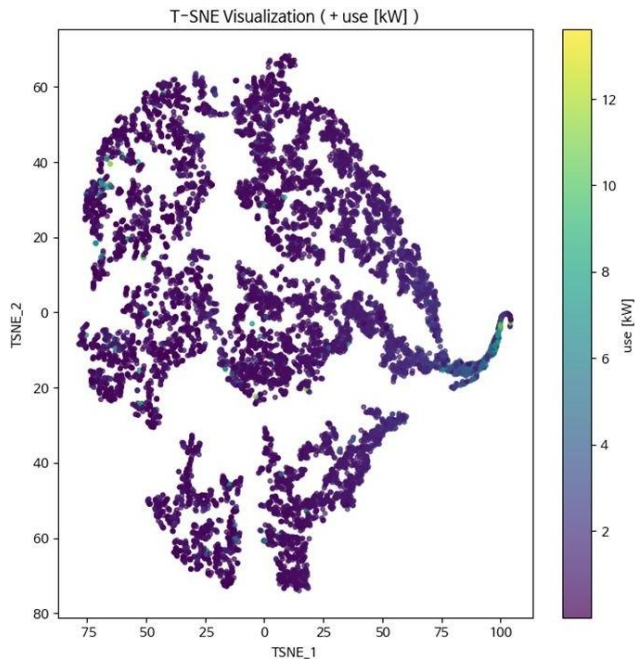
use [kW]와 House overall [kW]
은 같은 값을 갖고 있음

temperature의 높은 값들은
이상치로 처리하지 않음

● 변수 분포 및 변수 간 상관관계 확인



- T-SNE를 통한 변수 분포 확인



사용량에 대한 값이 특정 군집에 집중



특정 조건에서 유사한 에너지를 소비함

데이터 전처리

EDA 결과 반영

특정 변수 내 Dirty data 제거

결측값 제거

중복된 변수(house overall[kW]) 제거

특성 엔지니어링

시간 기반 특성:

- hour: 시간
- end: 주말 여부
- day_of_week: 요일
- month: 월
- is_week
- season: 계절

상호작용 특성:

- temp_humidity_interaction:
날씨와 에너지 소비의 상관관계를 반영

이동 평균 및 lag 특성:

- use_lag_1, use_lag_2:
과거 1단계, 2단계의 에너지 사용량
- use_rolling_mean_3: 3단계 이동 평균

변수 간 비율 및 차이 계산:

- use_gen_ratio: 에너지 사용 대비 발전 비율
- temp_pressure_diff: 온도와 기압 차이

범주형 변수 처리

1. One-Hot Encoding
+ Target Encoding :

인코딩 후 PCA 적용 여부에
따라 구분하여 저장

2. One-Hot Encoding
+ Hash Encoding:

인코딩 후 PCA 적용 여부에
따라 구분하여 저장

> PCA 과정: 전체 분산의 80% 이상을
설명하는 주성분을 선택

변수 선택

1. 변수를 하나씩 추가 및 제거하며 VIF를
계산, 최종적으로 최대한 많은 변수를 포
함하며 8 미만의 VIF 유지

2. Stepwise Selection 수행 결과가 미흡
하여 정성적 및 정량적 분석을 통해 기상
변수를 임의로 추가

데이터 모델링

- 모델 선정

모델	장점
Linear Regression	선형 관계를 잘 설명하고 직관적이며 계산이 빠름
Lasso	변수 선택과 모델 단순화가 가능하며, 과적합을 줄이는 데 효과적
Ridge	다중공선성 문제를 해결하는 데 유용
ElasticNet	Lasso와 Ridge의 장점을 모두 활용 가능
Bayesian Ridge	베이지안 추정을 사용하여 데이터의 불확실성을 모델링하고, 과적합을 방지

모델 학습 및 평가

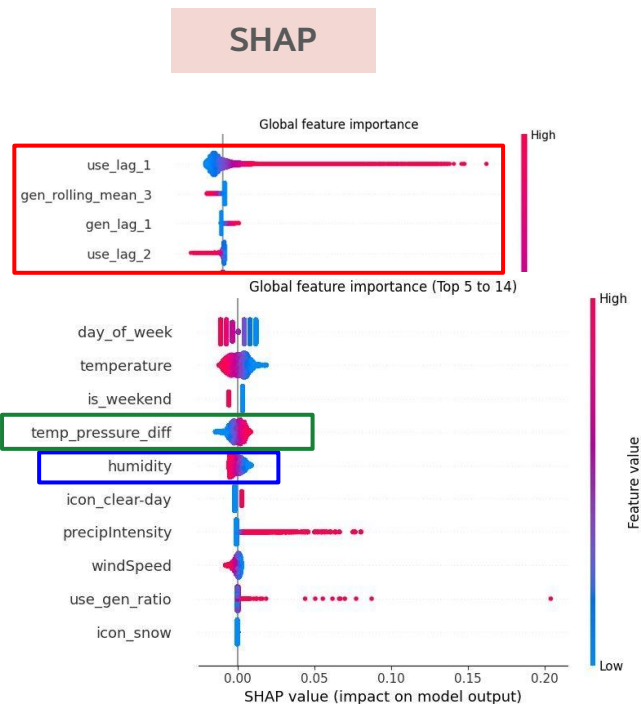
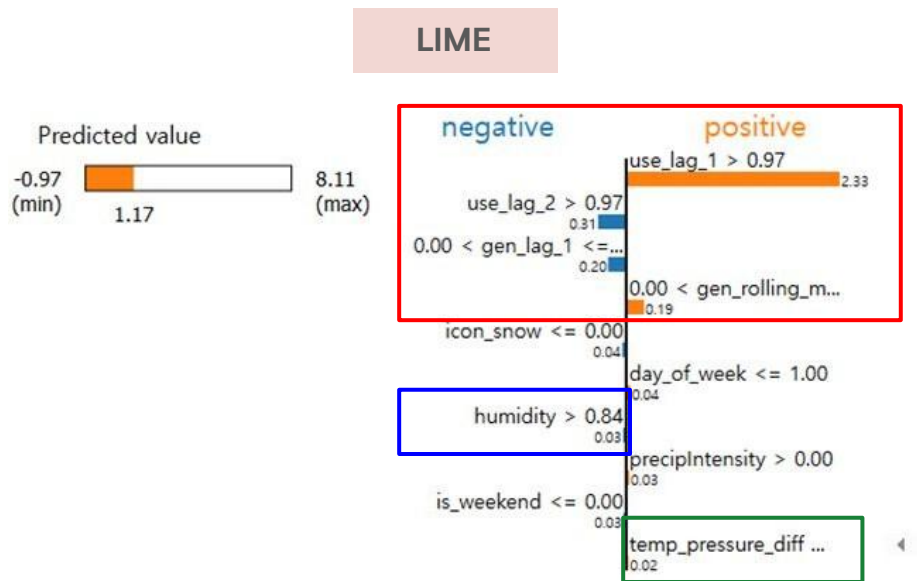
- $\langle \text{train} : \text{validation} : \text{test} = 70 : 15 : 15 \rangle$ 해당 비율로 데이터 분리하여 일반화 성능 향상
- 하이퍼파라미터 튜닝 \rightarrow 베이지안 최적화, k-fold
 - ➡ 베이지안 최적화: 이전 결과를 기반으로 하이퍼파라미터를 선택해 효율적으로 최적화
(기존 회귀 모델의 하이퍼파라미터 튜닝 전 결과보다 적은 비용으로 더 좋은 성능을 얻고자 시도)
 - ➡ k-fold: 모델이 훈련 데이터에 과적합(overfitting)되지 않고 새로운 데이터에서도 잘 동작하는지를 확인

모델 학습 및 평가

● 결과

전처리 방식			적용 모델		하이퍼파라미터 튜닝	MAE	MSE	MedAE	R2	Adj R2
원핫인코딩	타겟인코딩	PCA X	Linear Regression	X		0.1404	0.1703	0.0476	0.8469	0.7856
			Linear Regression	O		0.1404	0.1703	0.0476	0.8469	0.7856
			Lasso	X		0.5536	1.0324	0.4035	0.0718	-0.2994
			Lasso	O		0.1802	0.1855	0.0899	0.8332	0.7665
			Ridge	X		0.1404	0.1703	0.0476	0.8469	0.7856
			Ridge	O		0.1404	0.1703	0.0476	0.8469	0.7856
			ElasticNet	X		0.3876	0.5228	0.2736	0.53	0.342
			ElasticNet	O		0.143	0.1708	0.0496	0.8464	0.785
			BayesianRidge	X		0.1404	0.1703	0.0476	0.8469	0.7856
			BayesianRidge	O		0.1404	0.1703	0.0476	0.8469	0.7856
원핫인코딩	타겟인코딩	PCA O	Linear Regression	X		0.1777	0.2251	0.0556	0.7977	0.7639
			Linear Regression	O		0.1777	0.2251	0.0556	0.7977	0.7639
			Lasso	X		0.5753	1.1123	0.4198	0	-0.1667
			Lasso	O		0.2664	0.2652	0.1615	0.7615	0.7218
			Ridge	X		0.1777	0.2251	0.0556	0.7977	0.7639
			Ridge	O		0.1777	0.2251	0.0556	0.7977	0.7639
			ElasticNet	X		0.4471	0.6486	0.321	0.4169	0.3197
			ElasticNet	O		0.1789	0.225	0.0575	0.7977	0.764
			BayesianRidge	X		0.1777	0.2251	0.0556	0.7977	0.7639
			BayesianRidge	O		0.1777	0.2251	0.0556	0.7977	0.7639
원핫인코딩	해시인코딩	PCA X	Linear Regression	X		0.1404	0.1703	0.0476	0.8469	0.7478
			Linear Regression	O		0.1404	0.1703	0.0476	0.8469	0.7478
			Lasso	X		0.5536	1.0324	0.4035	0.0718	-0.5287
			Lasso	O		0.1802	0.1855	0.0899	0.8332	0.7253
			Ridge	X		0.1404	0.1703	0.0475	0.8469	0.7478
			Ridge	O		0.1404	0.1703	0.0475	0.8469	0.7478
			ElasticNet	X		0.3876	0.5228	0.2736	0.53	0.2259
			ElasticNet	O		0.143	0.1708	0.0496	0.8464	0.7471
			BayesianRidge	X		0.1404	0.1703	0.0475	0.8469	0.7478
			BayesianRidge	O		0.1404	0.1703	0.0475	0.8469	0.7478
원핫인코딩	해시인코딩	PCA O	Linear Regression	X		0.1777	0.2251	0.0556	0.7977	0.7639
			Linear Regression	O		0.1777	0.2251	0.0556	0.7977	0.7639
			Lasso	X		0.5753	1.1123	0.4198	0	-0.1667
			Lasso	O		0.2664	0.2652	0.1615	0.7615	0.7218
			Ridge	X		0.1777	0.2251	0.0556	0.7977	0.7639
			Ridge	O		0.1777	0.2251	0.0556	0.7977	0.7639
			ElasticNet	X		0.4471	0.6486	0.321	0.4169	0.3197
			ElasticNet	O		0.1789	0.225	0.0575	0.7977	0.764
			BayesianRidge	X		0.1777	0.2251	0.0556	0.7977	0.7639
			BayesianRidge	O		0.1777	0.2251	0.0556	0.7977	0.7639

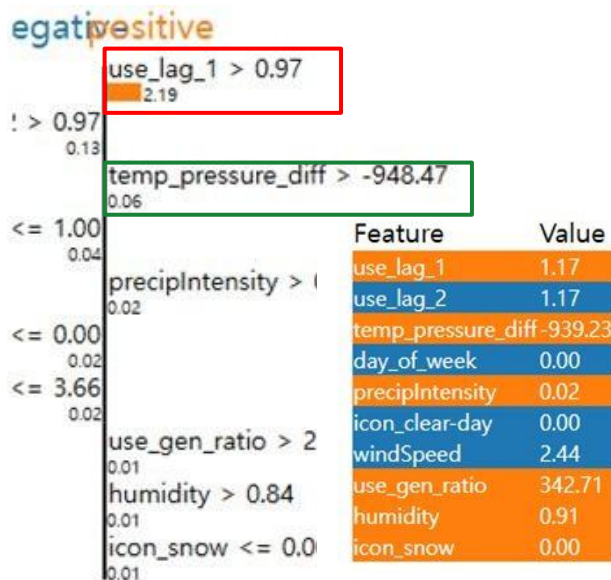
Linear Regression



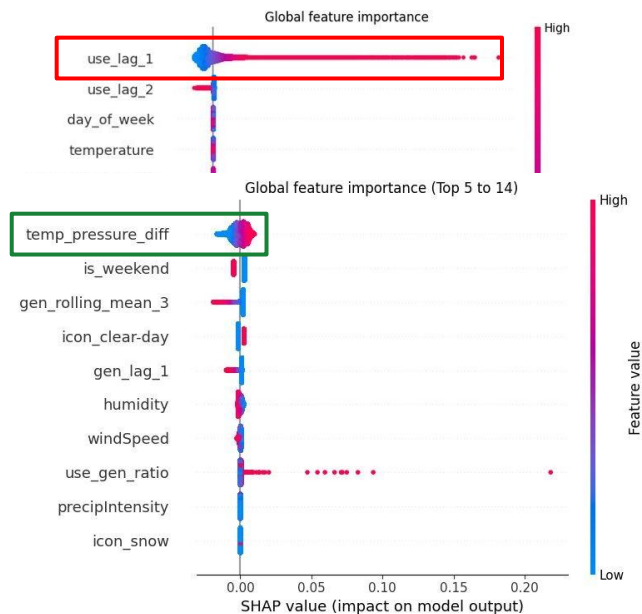
결과 해석

● ElasticNet

LIME



SHAP



● 기대효과

맞춤형 에너지 절약 솔루션 개발

예측 모델을 활용하여
가구별 에너지 소비 패턴 분석



개인화된 에너지 절약 방안을 제안



날씨에 따른 에너지 사용 계획을
자동화하는 맞춤형 서비스 제공

효율적인 전력 생산 및 배분

전력 소비량 예측



재생에너지와 같은 변동성 높은 전력원도
안정적으로 활용할 수 있음



전력 예비율 최적화 및 불필요한
전력 생산을 줄이는 데 기여

지속 가능한 환경 조성

날씨와 에너지 소비간 관계 분석



에너지 효율성 증가



지역별 특화 에너지 정책 수립 등
기후 변화 대응 전략 수립에 기여

● 한계

제한된 데이터 문제

데이터가 특정 시간 및 장소에 국한됨



기상정보를 온전히 데이터에
반영하는데 어려움

시계열 특성 중 하나가 모델 예측에 큰 영향을 미침

모델이 해당 변수에 크게 의존



다른 잠재적인 변수나 환경적 요인을
충분히 반영하지 못했을 가능성 있음

외부 요인 반영의 한계

외부 요인이 실제 에너지 소비에 영향을
주어 예측 정확도에 반영될 수 있음



날씨 및 시간 변수에 중점을 두어
경제 상황, 에너지 가격, 지역별 인프라와
같은 외부 요인을 반영하지 못함

수행계획

● 일정 및 역할



Project Manager	송승규	모델링	강민혁, 이은강, 이르판	PPT	이주희
데이터 EDA 및 전처리	송승규, 강민혁	발표 (영상)	송승규		
특성 엔지니어링	문하늘	선행연구조사&관련자료	강민혁, 문하늘		

감사합니다.

Reference

1. Mitali, J. (2022). Energy storage systems: a review. *Energy Storage and Saving*, 1(3), 166-216. 10.1016/j.enss.2022.07.002
2. Rahul, S. (2023). Progress and challenges in electrochemical energy storage devices: Fabrication, electrode material, and economic aspects. *Chemical Engineering Journal*, 468. 10.1016/j.cej.2023.143706
3. Huang, Y. (n.d.). Key Challenges for Grid-Scale Lithium-Ion Battery Energy Storage. *Advanced Energy Materials*, 12(48). 10.1002/aenm.202202197
4. Dumont, O. (2020). Carnot battery technology: A state-of-the-art review. *Journal of Energy Storage*, 32. 10.1016/j.est.2020.101756
5. Götz, M. (2016). Renewable Power-to-Gas: A technological and economic review. *Renewable Energy*, 85, 1371-1390. 10.1016/j.renene.2015.07.066
6. 김철현. (2013). 일별·시간대별 전력 부하패턴 분석 [KEEI 수시연구보고서 13-03].
<https://www.keei.re.kr/library/10310/contents/6842397>
7. Tuohy, A. (2019, Apr 29). Forecasting Reserves: Current Research, a New Tool, and Next Steps. ESIG. Retrieved Nov 29, 2024, from <https://www.esig.energy/forecasting-reserves-current-research-a-new-tool-and-next-steps/>
8. https://www.korea.kr/briefing/pressReleaseView.do?newsId=156547521&utm_source=chatgpt.com