

세상에서 가장 가벼운 단기 전력 수요 예측 모델

나인성 | 남영 | 박지호 | 안태진 | 황승주

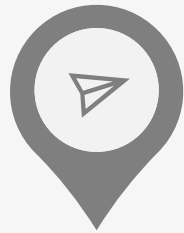
세상에서 가장 가벼운 단기 전력 수요 예측 모델

Index



프로젝트 소개

- i. 문제 정의 가설 도출
- ii. 문제 해결 방법 정의



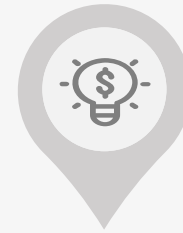
데이터 전처리

- i. 데이터 수집
- ii. 데이터 보간 (Interpolate)



EDA & Insight

- i. 전력 수요 데이터
- ii. 기상 데이터
- iii. Insight 도출



가설 검증

- i. 모델 구축
- ii. 요인 비교



결론 및 Insight

- i. 결론 및 의의
- ii. 부록

기상정보 없이 전력 수요 예측이 가능하다면?

- 일반적인 분석 방법

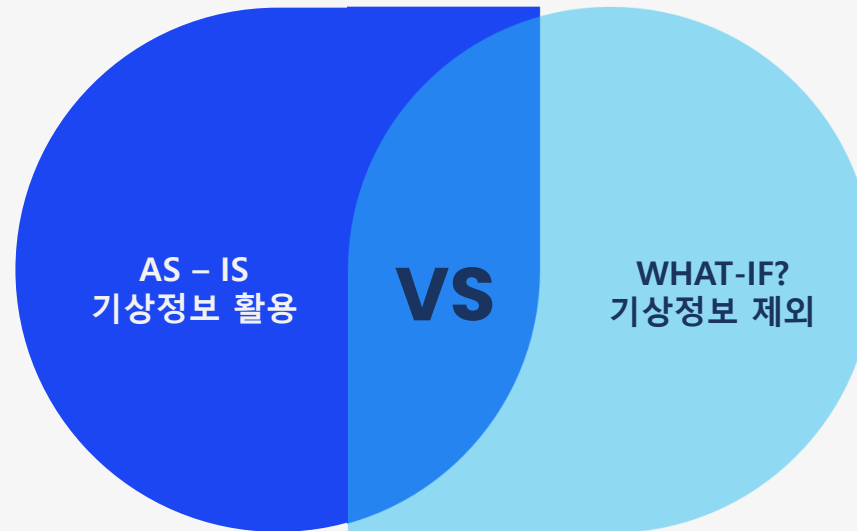
- 전력 수요 예측에 기상정보를 활용하는 것이 일반적인 방법

- 방대한 데이터의 양

- 방대한 초단기 기상데이터를 다룸

- 다수의 결측치 존재

- 기기의 고장/운영중단이 야기하는 잦은 결측 이슈



- 적은 데이터 양

- 추가 정보 없는 전력 수요 예측이 가능해짐

- 노이즈 영향 축소

- 기상 데이터 제외 시, 결측치로 인한 노이즈의 영향 축소

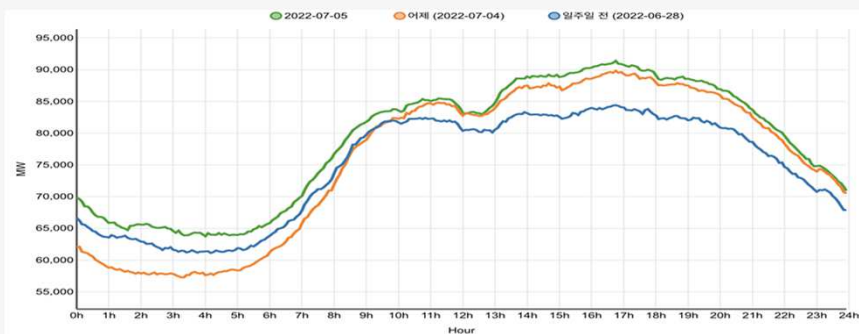
- 경제성

- 보다 적은 비용으로 단기 전력수요 예측 가능

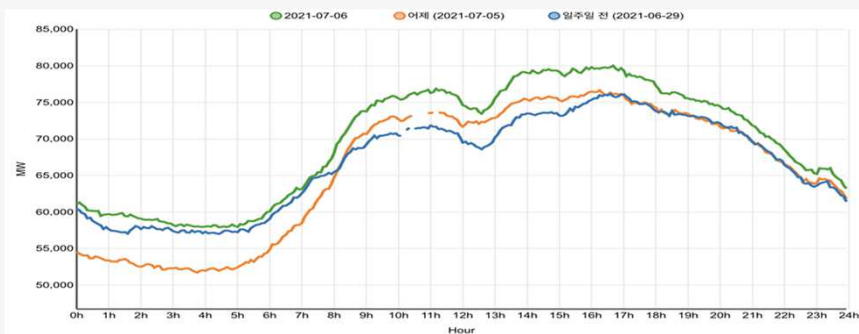
프로젝트 소개

문제정의 가설 도출

2021.07 1st week Tuesday



2021.07 1st week Wednesday



문제 정의 & 가설 도출



과거의 패턴만으로 미래의 전력 수요 예측이 가능하다.

기상 정보를 활용한 모델과의 MAPE 차이는
1%p 이상 발생하지 않을 것이다.

외부요인보다는 과거 수요와 최근 방향성의 트렌드가
미래의 전력수요를 예측하는 중요한 변수일 것이다.

전력 수요는 일정한 패턴을 보이는 시계열 데이터이다.

문제 해결 절차 정의

1 데이터 수집 및 크롤링

2 데이터 결측치 보간 및 전처리

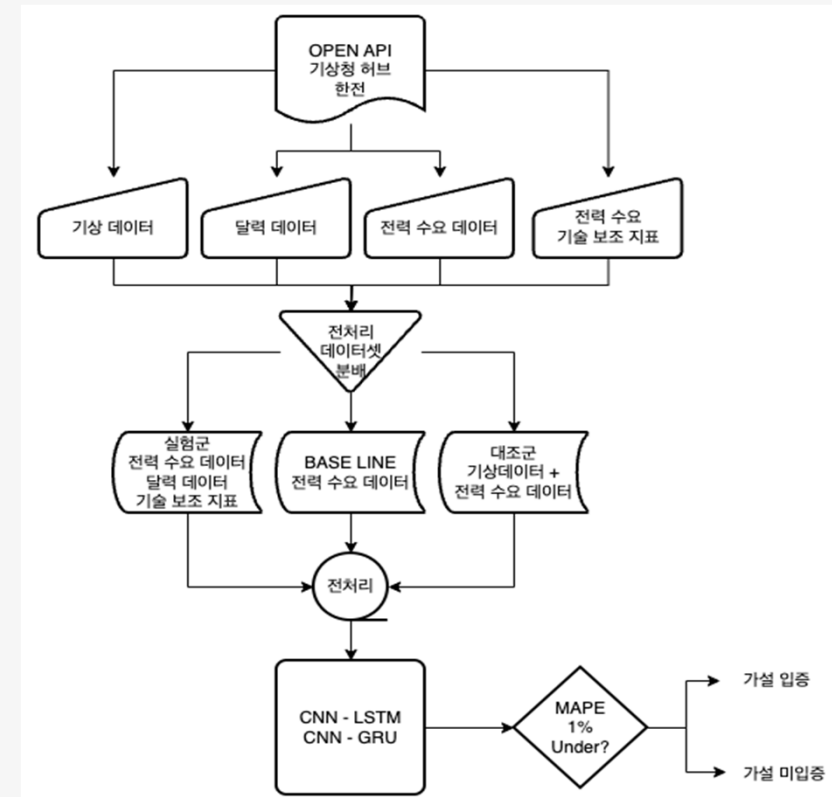
3 신규 변수 생성 (기술보조지표, 달력데이터)

4 실험 및 대조군 데이터셋 분리

5 모델 구축

6 동일 환경에서 실험 진행

7 평가



데이터 수집 및 전처리 ⚡

데이터 수집 | 데이터 보간

데이터 수집

종류	특징	AS-IS	TO-BE
기상 데이터 종관 기상 관측 (ASOS)	<ul style="list-style-type: none"> 정해진 시각의 대기 상태를 파악하기 위해 모든 관측소에서 같은 시각에 실시하는 지상관측 시정, 구름, 증발량, 일기현상 등 일부 목적 요소를 제외하고 자동으로 관측 	<ul style="list-style-type: none"> 방대한 데이터 수집 및 전처리 필요 지역, 관측장비, 기상인자별 불규칙적인 결측 값 처리 필요 전력수요 데이터와의 통합을 위한 기상 대표 값 도출 필요 	<ul style="list-style-type: none"> 크롤링 및 parquet 파일 처리 보간법, 대표 값 도출을 통한 값 대체 전국 주요 광역시 및 대표 도시 선정하여 대표 값으로 활용
전력수요 데이터	<ul style="list-style-type: none"> 실시간 전력수요 현황 집계 데이터 	<ul style="list-style-type: none"> 방대한 데이터 수집 및 전처리 필요 불규칙적인 기간의 결측 값 처리 필요 	<ul style="list-style-type: none"> 크롤링 및 parquet 파일 처리 보간법, 대표 값 도출을 통한 값 대체
달력 데이터	<ul style="list-style-type: none"> 분석 기간 내 공휴일 정보 	<ul style="list-style-type: none"> sparse 데이터 처리 필요 	<ul style="list-style-type: none"> 원핫 인코딩 활용

AS - IS

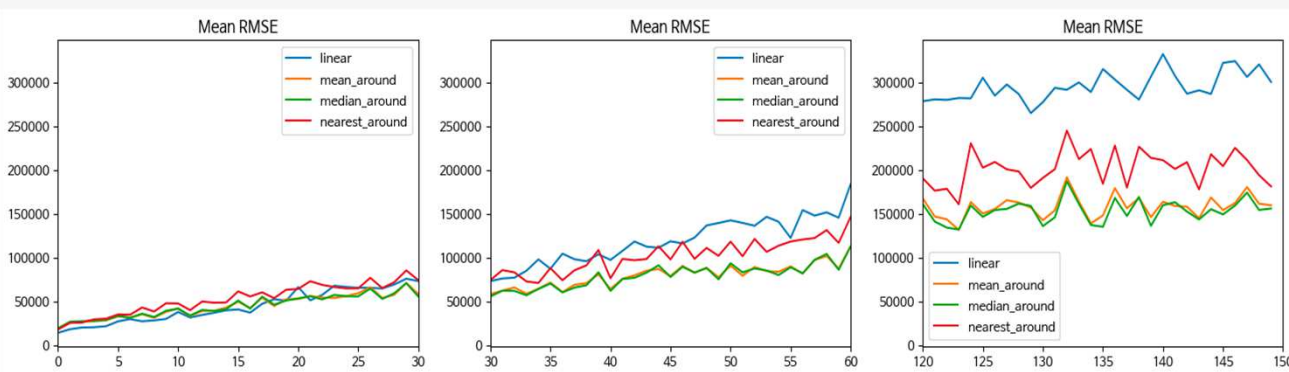
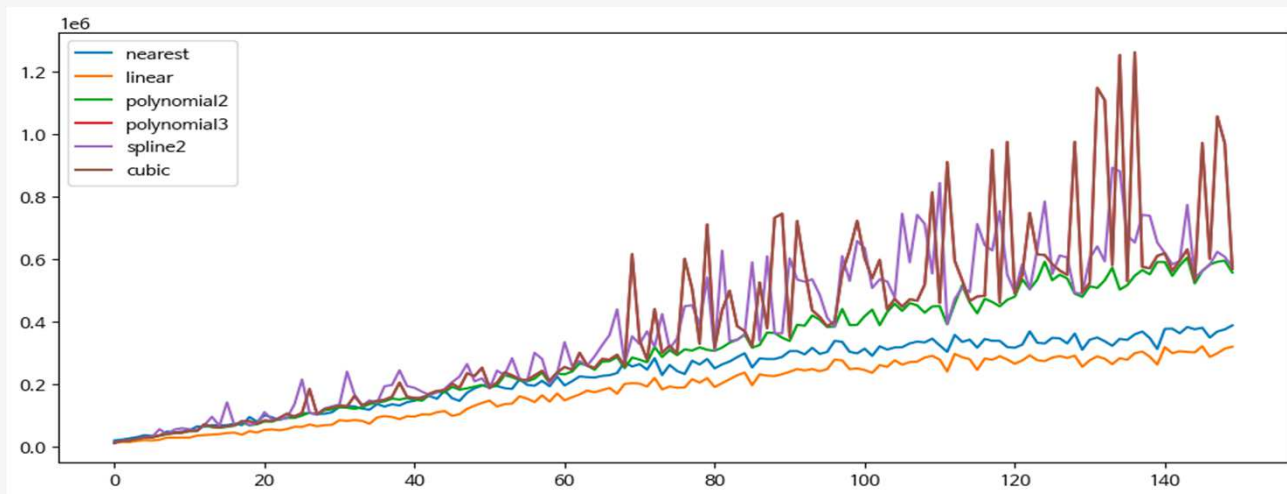
- 3년간의 5분 간격의 방대한 기상, 전력 수요 데이터 수집/처리
- 불규칙하며 연속적인 결측 값. 많은 데이터의 처리 필요
- 비연속적인 요일, 시간, 휴일, 공휴일 데이터의 변수처리 필요

TO - BE

- 데이터 수집 자동화 및 parquet 파일 처리
- 기성 보간법 활용 및 유사 패턴의 인근 일자를 활용한 보간
- 원핫인코딩 활용

데이터 수집 & 전처리

최적의 보간법 도출



단기 결측 구간 실험 "Linear is the best"

- Nearest : 최근접 이웃 보간법
- Linear : 선형 보간법
- Polynomial2 : 2차 다항식 보간법
- Polynomial3 : 3차 다항식 보간법
- Spline : 스플라인 보간법
- Cubic : 3차 보간법

장기 결측 구간 실험 "Median is the best"

- Mean : 주변 패턴의 평균값으로 보간
- Median : 주변 패턴의 중위 값으로 보간
- Grid : 주변 패턴 중 RSME가 가장 낮은 패턴으로 보간

INSIGHT. 결측 패턴 별 보간 규칙 도출 '20일 전후'

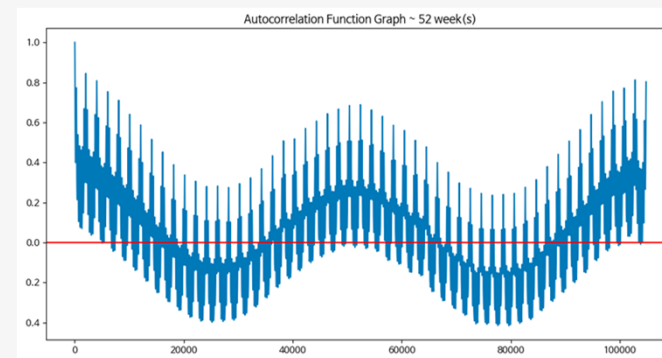
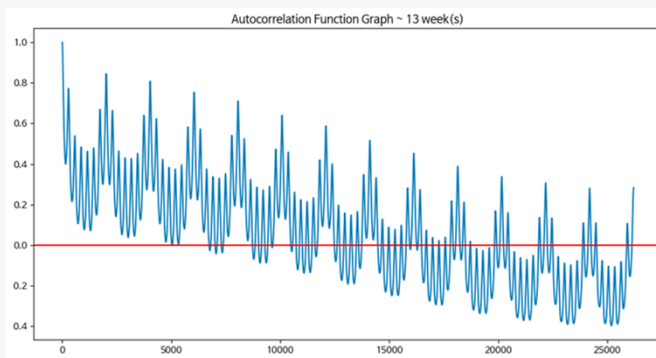
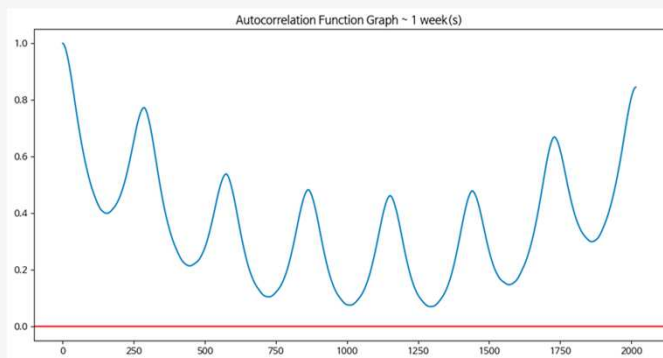
- 연속 20일 이상의 결측 구간
- 중앙값으로 보간법 실행
- 연속 20일 미만의 결측 구간
- 선형 보간법 실행

EDA & INSIGHT

전력수요 데이터 EDA | 기상 데이터 EDA | Insight Wrap Up

데이터 수집 & 전처리

전력수요 데이터의 자기상관성 확인



FACT

뚜렷한 계절성 확인

- 1년간 상승, 하강 패턴이 4계절 기온의 방향성과 일치
- 1계절 내 파동의 진폭 패턴이 일정한 간격을 유지
- 1주중 내 파동의 진폭 패턴이 일정한 간격을 유지

FACT

동일 요일, 시간의 영향도 높음

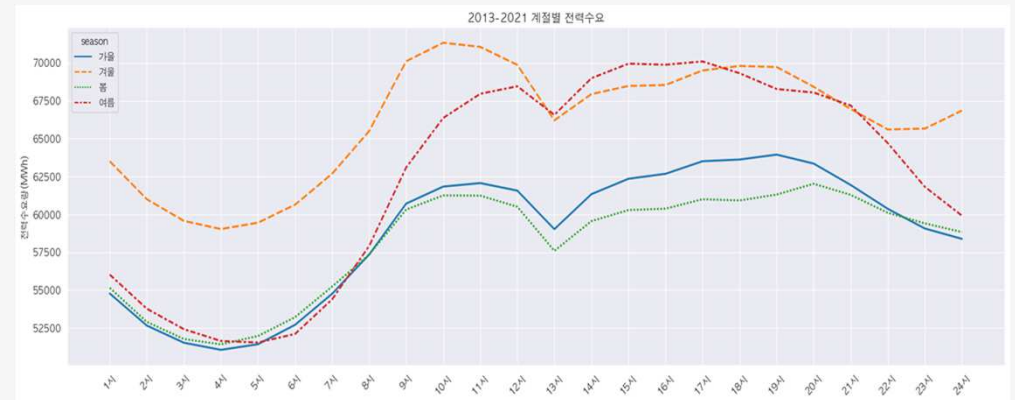
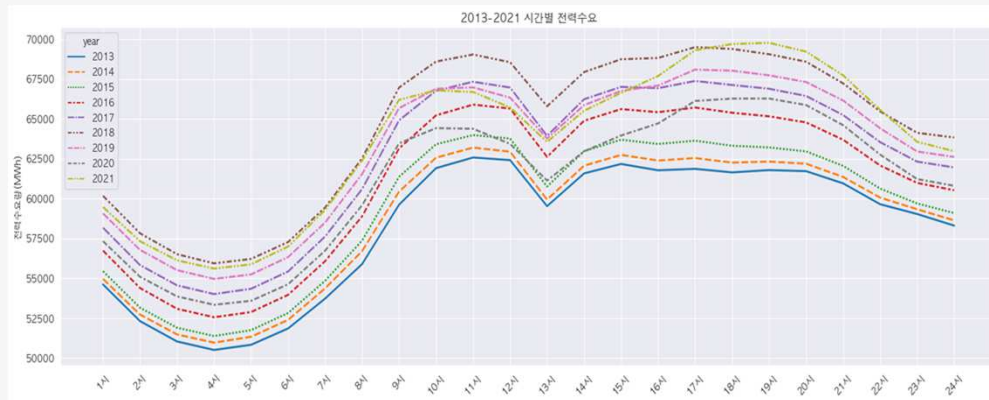
- 12시간 하강, 12시간 상승하는 일정한 패턴
- 24시간 이후의 자기상관성보다 1주일 이후의 자기상관성이 높은 패턴
- 시간이 지날수록 위 패턴으로 진동하며 자기상관성이 하강

INSIGHT

과거 패턴이 예측에 가장 유용한 변수일 것

- 12시간 하강, 12시간 상승하는 일정한 패턴
- 24시간 이후의 자기상관성보다 1주일 이후의 자기상관성이 높은 패턴
- 시간이 지날수록 위 패턴으로 진동하며 자기상관성이 하강

전력수요 데이터의 시간, 계절 변수의 중요성 확인



FACT

시간대 별 명확한 패턴 확인

- 11, 18-19시경, 전력사용량 피크
- 전체 데이터 구간(9년)간 큰 변화 없는 안정적인 패턴

FACT

계절간 패턴의 특징 확인

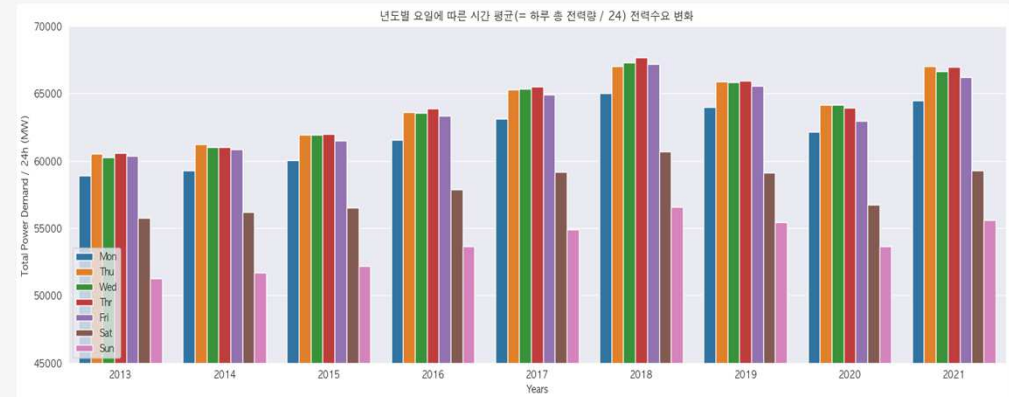
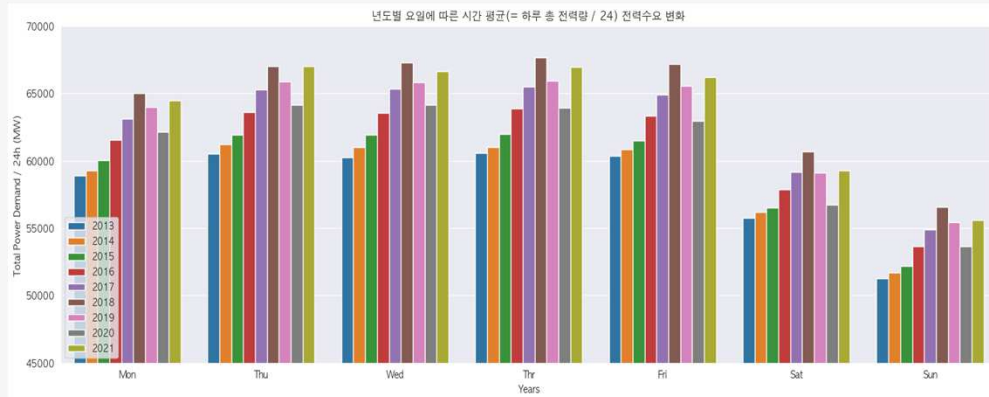
- 여름, 겨울의 경우 에너지 수요량의 변동폭이 크다
- 여름, 겨울의 경우 시간대별 기온변화의 영향이 크다
- 상대적으로 봄, 가을의 전력수요 데이터가 안정적이다

INSIGHT

시간, 계절 변수가 중요한 예측 변수일 것

- 시간, 계절을 구분하는 변수의 중요성이 크다

전력수요 데이터의 요일변수의 중요성 확인



FACT

요일별 전력수요의 명확한 패턴 확인

- 전체 데이터 구간(9년)간 전력수요 패턴 유지
- 화-금, 월, 토, 일 순의 전력수요 패턴

FACT

사회 / 계절 이슈가 반영됨을 확인

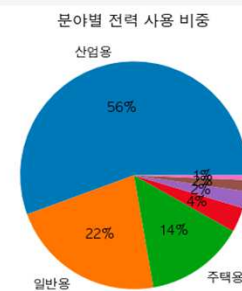
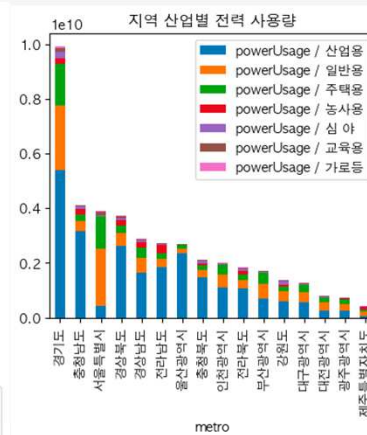
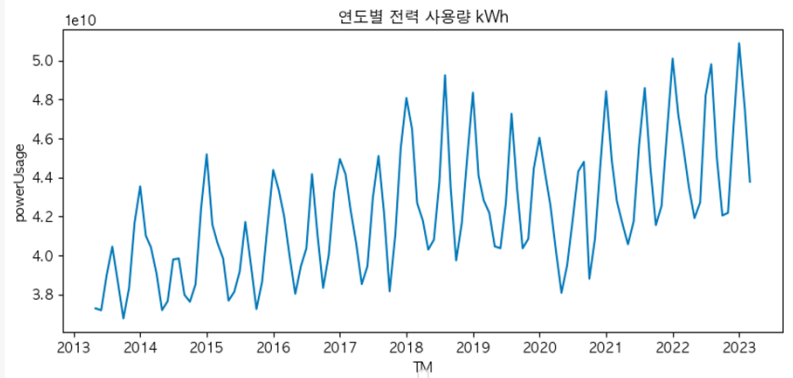
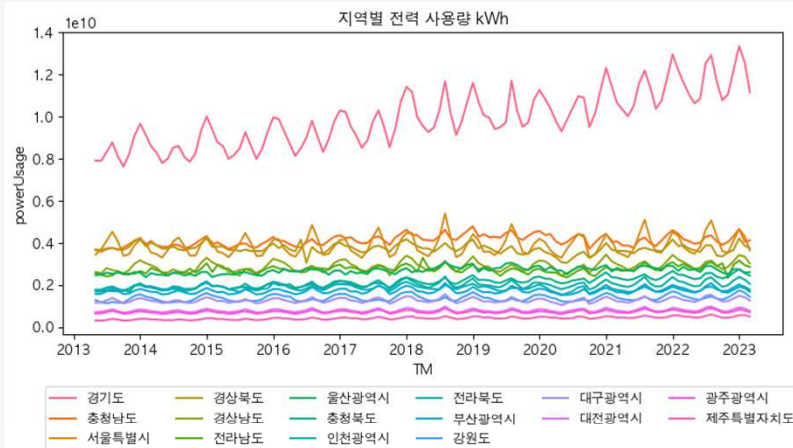
- 2018년의 수요 급상승
 - 2018년 여름 기록적 폭염의 영향
- 2019-2020년의 추세 역행
 - 코로나19의 영향
- 2013-2021년
 - 큰 추세상 수요 상승세 지속

INSIGHT

요일 변수가 중요한 예측 변수 일 것

- 요일을 구분하는 변수의 중요성이 크다
- 요일 패턴을 보아, 업무 및 산업 수요가 주요 수요처일 것으로 추정
- 사회 / 계절 이슈가 반영되어 있으므로, 불완전한 정보를 변수로 추가할 필요성이 낮다

전력수요 데이터의 지역 / 산업별 특징



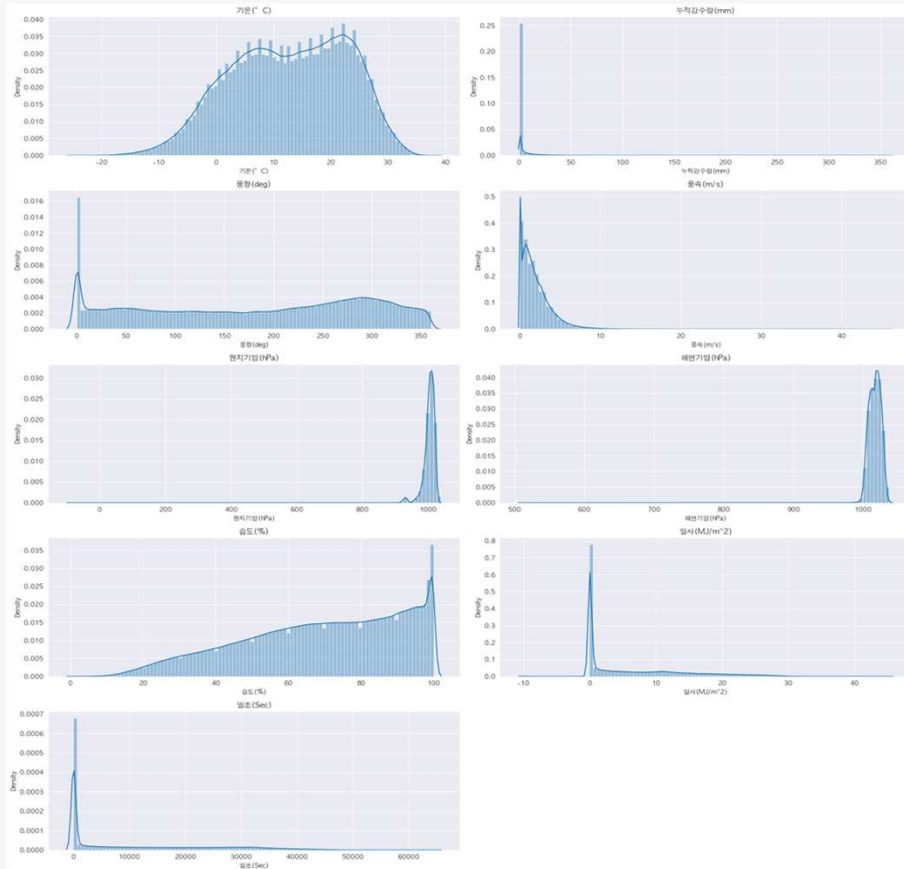
FACT

- 산업용 수요가 총 비중의 과반수 이상
- 총 전력수요는 지속적으로 상승
- 경기도의 성장세가 가파르게 상승
- 수년간 지역별 수요 순위 변화 없음
- 사회 / 계절 이슈가 수요에 반영됨

INSIGHT

- 산업 수요가 주요 수요처임
- 산업수요가 안정적인 수요패턴을 형성하는 이유로 추정
- 사회 / 계절 이슈가 반영되어 있음

기상데이터 특성



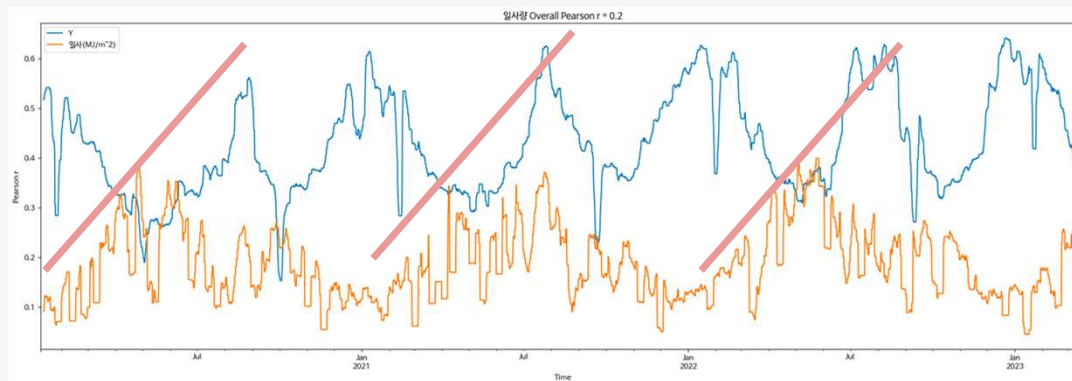
기상인자	결측 수	결측 비율	특징
기온	35533	14%	상대적으로 결측과 편향이 적음
누적 강수량	425443	21%	결측 유무의 판단에 어려움
풍향	68578	25%	특정한 패턴을 형성하지 않음
풍속	70622	41%	편향된 데이터
현지기압	37015	42%	편향된 데이터
해면기압	36769	22%	편향된 데이터
습도	37786	22%	편향된 데이터
일사	713033	42%	지역별 결측 비중 차이가 큼
일조	64176	38%	지역별 결측 비중 차이가 큼

INSIGHT

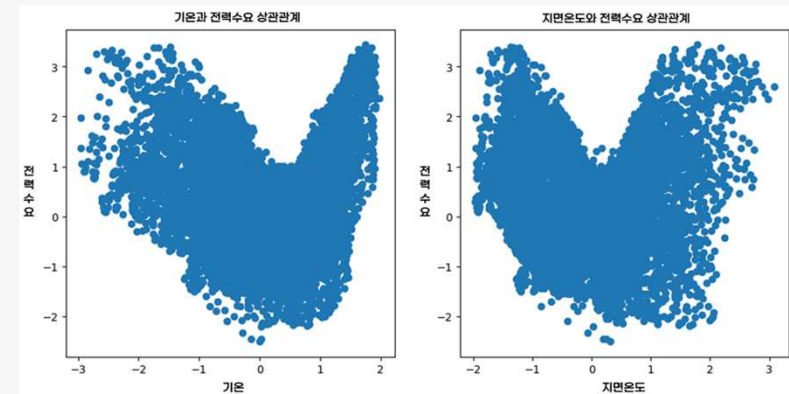
- 모든 변수에 결측치가 다수 존재하므로, 결측치 제거의 중요성을 확인

기상요인의 주요 예측 변수여부 확인

최근 3년 기상인자와 전력수요의 1주 이동 중앙값



최근 3년 기상인자와 전력수요의 상관관계



WE FOUND

- 기온, 전력수요와의 유사한 계절성
- 일사량, 기온 대비 1.7배의 상관관계
- 온도인자, 양극단에서 높은 상관관계

BUT

- 단기 / 초단기 패턴상 노이즈가 많다

SO INSIGHT

- 전력수요 예측에 기상인자의 영향력이 낮을 것으로 예상

Insight Wrap up

Heading	Details
전력수요 EDA 인사이트	<ul style="list-style-type: none"> 전력 수요는 명확한 계절성을 가지며, 연도에 관계없이 시간대별 패턴이 유지됨. 전력 수요에는 폭염이나 코로나와 같은 외부 요인이 반영되어 있으며, 계절성 패턴이 안정적으로 유지됨. 전력 수요는 대부분 산업 수요에 의해 영향을 받음. ACF 분석 결과 자기회귀성이 높은 일정한 패턴을 보이는 시계열 데이터임을 확인.
기상정보 EDA 인사이트	<ul style="list-style-type: none"> 전력 수요와 기상정보 간에는 상관관계가 있지만 노이즈가 많음. 온도와 일사량이 전력 수요와 상관관계가 있으며, 특히 온도의 양극단에서 높은 상관관계를 보임. 기상요인 중 전력 수요와의 상관관계가 높은 일사량 데이터의 결측이 많이 발생.
Wrap Up	<ul style="list-style-type: none"> 기상 인자는 전력수요에 일정 수준의 영향을 미침. 하지만 전력수요는 시간과 계절에 강건한 계절성을 보이므로, 이미 과거 패턴에 기상 이변 등이 반영되어 있을 것으로 예상됨. 따라서, 전력 수요 예측에 있어서 가장 중요한 변수는 과거의 패턴과 트렌드일 것. 과거 패턴의 영향력을 확인했으므로, 기술적 분석이 유효할 것으로 예상됨. 또한, 과거의 시간대와 요일 및 휴일 유무의 정보 역시 중요한 변수로 작용함을 확인함.
So	<ul style="list-style-type: none"> '과거의 패턴만으로도 미래의 전력 수요 예측이 가능하다.'는 가설 유지

가설 검증 & 결론

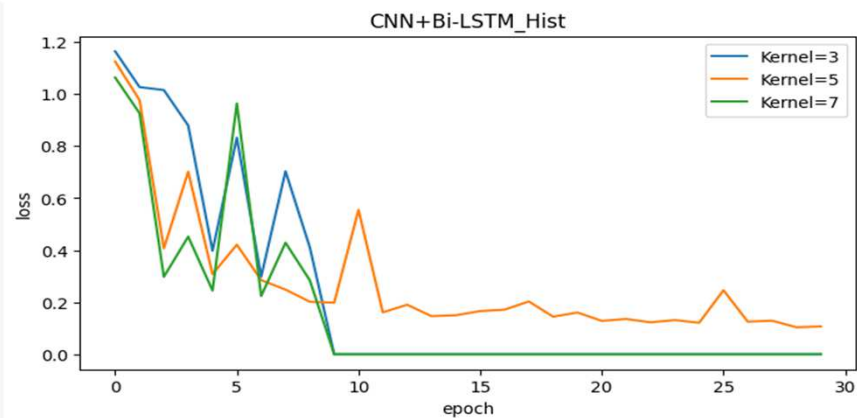
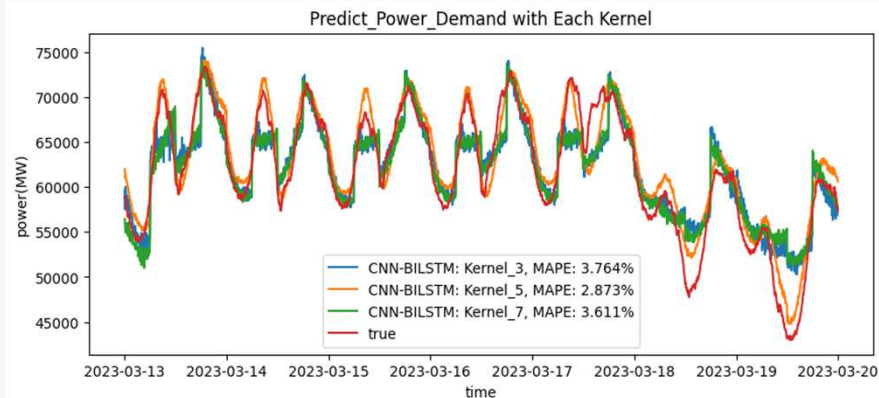
모델 구축 | 요인 비교 | 결론

가설 검증

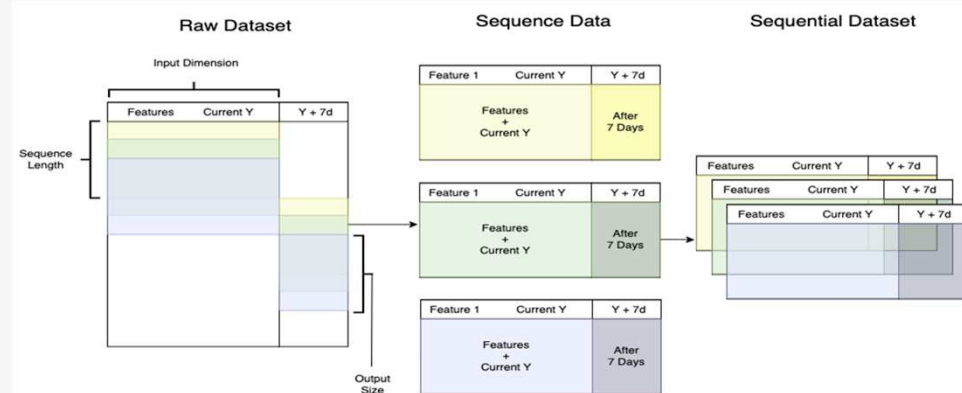
모델 구축

CNN Conv1d Kernel size 선정

kernel_size=5



Dataset 구조

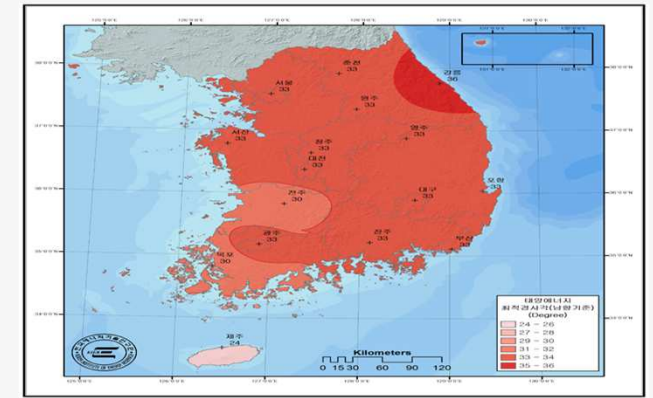
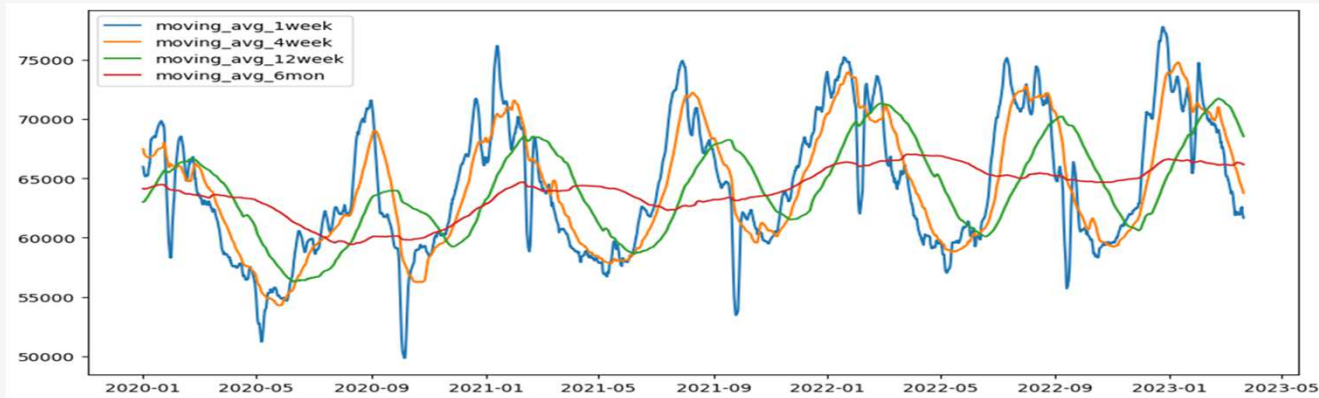


LSTM / GRU Parameters

- Sequence Length / Output Size : 자기상관성 최대치, 7일
- Input Dimension : 비교군 마다 상이
- Num of Layers : 3 (Grid Search result)
- Hidden Dimension : 288 (Grid Search result)

가설 검증

지표 및 대표 값 선정



기술 지표 선정

- 1일, 1주 전 동시간대 데이터
- 24시간 내 최고 전력 수요량
- 24시간 내 최저전력 수요량
- 6시간 이동평균 (30분 - 26주)

달력 데이터 선정

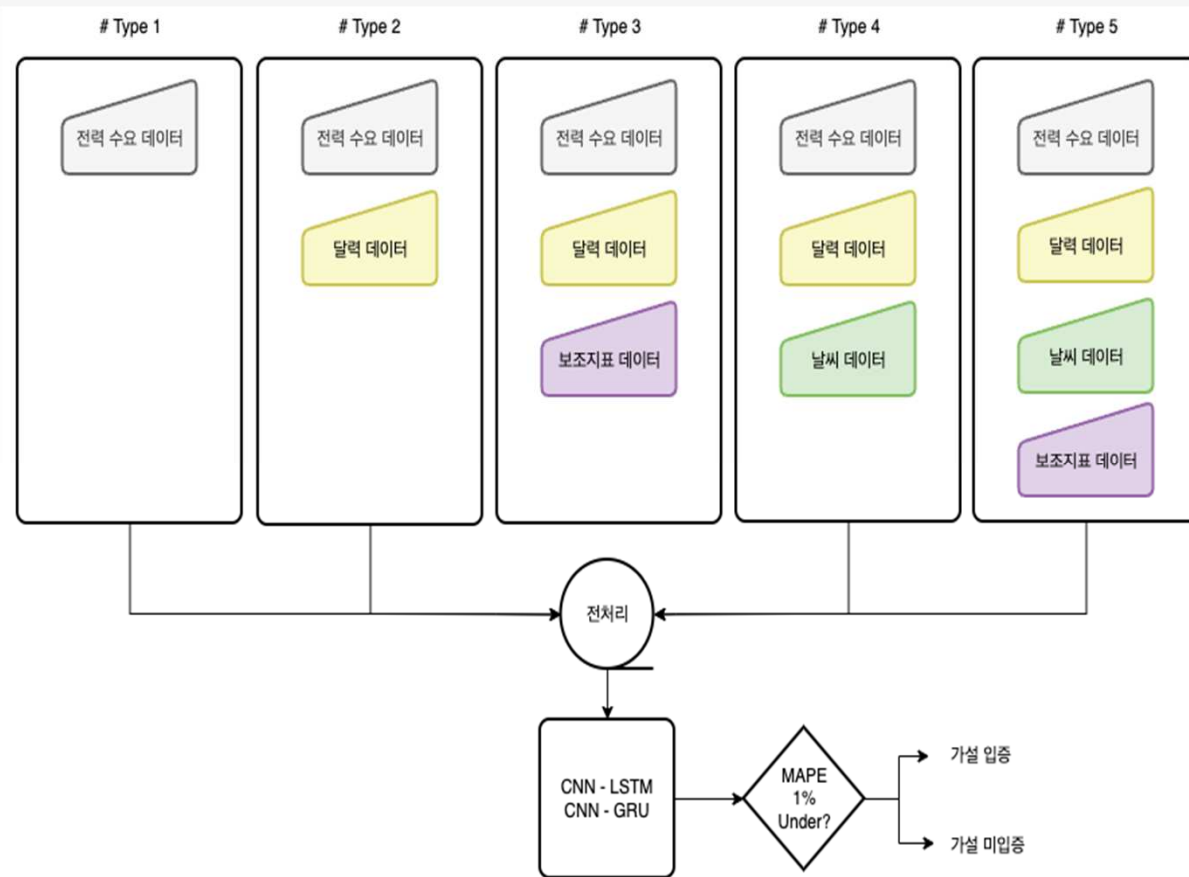
- 기준일자 3년간의 휴일, 공휴일, 요일 및 시간 추출
- 원핫 인코딩을 활용한 더미 변수 생성

기상 대표 값 선정

- 지역 가중치, 주요 특별시를 활용한 Grid Search 결과 비교
 - 주요 특별시의 정보를 활용한 기상 대표 값 선정 모델의 성능이 우수
 - 해당 기준을 활용하여 기상 대표 값 선정

가설 검증

실험군 및 비교군 선정



조건

- 입력 데이터의 종류 외 랜덤시드, 하이퍼 파라미터, 스케일링 등 다른 조건은 모두 동일하게 설정한다.

가설 입증 조건

- MAPE of Type2 and 3 \leq Type 4**
- Type2, Type3 (기상데이터 X) Type.4 (기상데이터 O)

데이터셋 준비

- Train
 - 2020.01.01 00:00:00 – 2023.02.05 23:55:00
- Validation
 - 2023.02.06 00:00:00 – 2023.03.06 23:55:00
- Test
 - 2023.03.13 00:00:00 – 2023.03.19 23:55:00

가설 검증

실험결과 및 비교

• Type2, Type3 (실험군)이 Type4 (대조군)의 성능보다 우수하다. (Fig. 3)

파라미터	# Type 1	# Type 2	# Type 3	# Type 4	# Type 5
피쳐	전력 수요	전력 수요 달력	전력 수요 달력 전력 보조지표	전력 수요 달력 날씨	전력 수요 달력 날씨 전력 보조지표
입력 컬럼 개수	1	55	66	58	69
합성곱 필터로 생성될 out의 Dimension	입력 컬럼 개수 * 2	입력 컬럼 개수 * 2	입력 컬럼 개수 * 2	입력 컬럼 개수 * 2	입력 컬럼 개수 * 2
합성곱 필터의 커널의 크기	5	5	5	5	5
은닉 상태의 개수	288	288	288	288	288
LSTM / GRU Layer 개수	3	3	3	3	3
Learning Rate	0.0001	0.0001	0.0001	0.0001	0.0001
Epochs	50	50	50	50	50
MAPE - CNN-LSTM(bi)	13.09%	3.29%	3.26%	3.50%	2.88%
MAPE - CNN-GRU(bi)	8.75%	3.38%	3.48%	3.80%	4.78%

결과 분석 및 가설 입증

실험 결과 분석

MAPE 0.05%p 이하의 차이는 동일한 성능으로 간주함.

- A. Type 1을 제외한 모든 모델의 성능이 KPX(한국전력거래소 예측치) 대비 7%p 이상 높다. (fig. 6)
 - i. → 전력데이터와 달력데이터만을 가지고 기존 모델을 대체할 수 있다.
- B. Type 2, Type 4의 성능은 각각 3.29%, 3.50% 로, Type 2의 성능이 Type 4 대비 약 6% 높다.(fig. 3)
 - i. → 날씨 데이터의 추가는 성능 저하로 나타난다.
- C. Type 2, Type 3의 성능은 각각 3.29%, 3.26% Type 2와 Type 3의 성능은 동등하게 관찰 된다. (fig. 4)
 - i. → 달력데이터와 보조지표의 조합은 성능개선이 미미하다.
 - ii. → 성능이 유사하다면, 경제성의 원리에 따라 Type 2 모델이 적합하다.
- D. Type 5 케이스를 볼 때, 모든 지표를 사용한다면 (전력 수요, 달력, 날씨, 전력 보조지표) 최대 2.88% 까지의 성능 개선이 가능하다.(fig. 5)
 - I. 단, 모델 별 편차를 볼 때, Type 2,3가 보다 강건함을 확인할 수 있다.

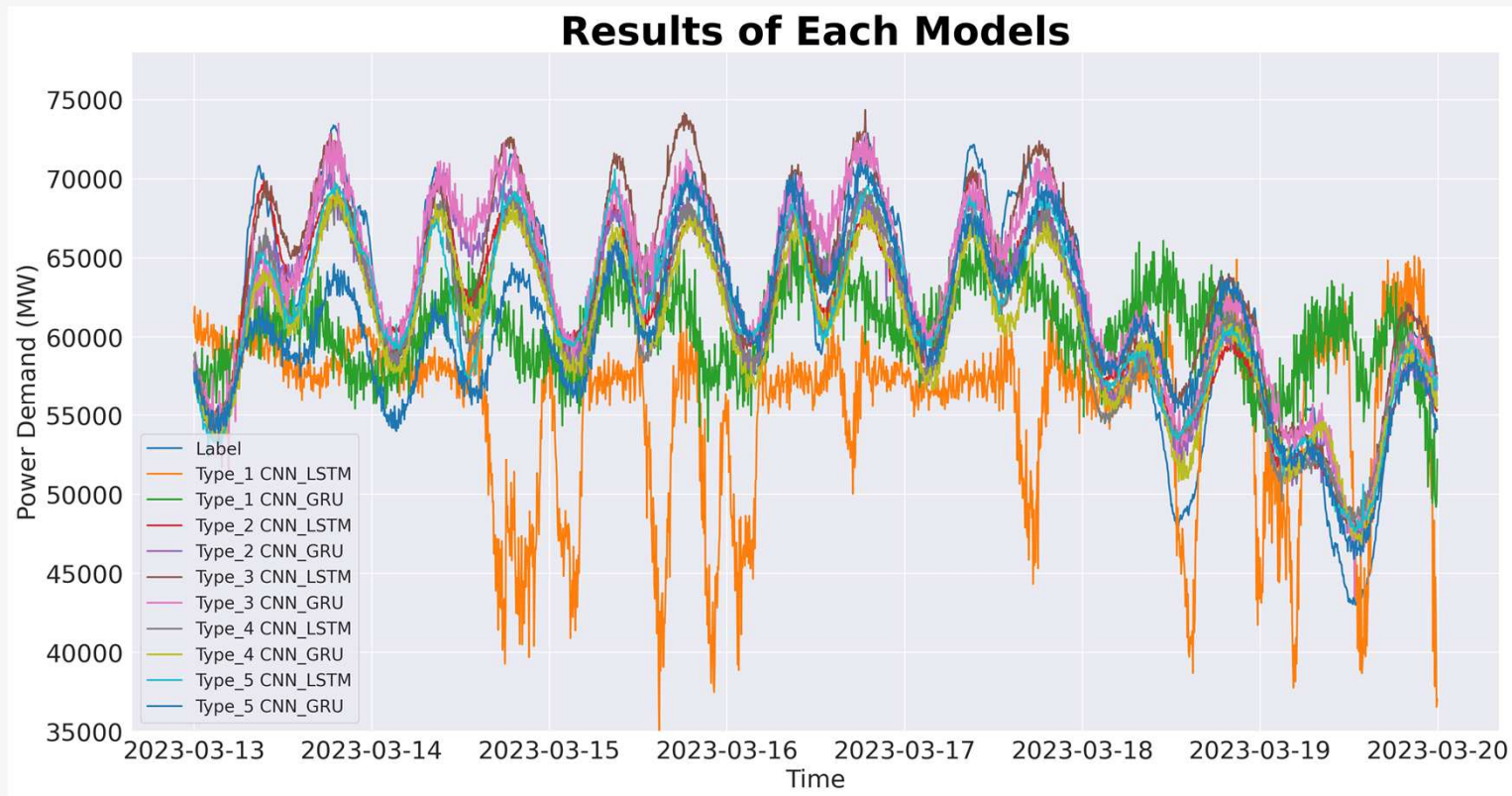
결론 도출

- A. Type 2의 입력인 전력수요, 달력 데이터 만으로도
기상정보를 사용하는 것과 동등한 성능이 발현됨을 검증함
- B. 따라서 '과거의 패턴만으로도 미래의 전력 수요 예측이 가능하다.'는 참이다.

Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

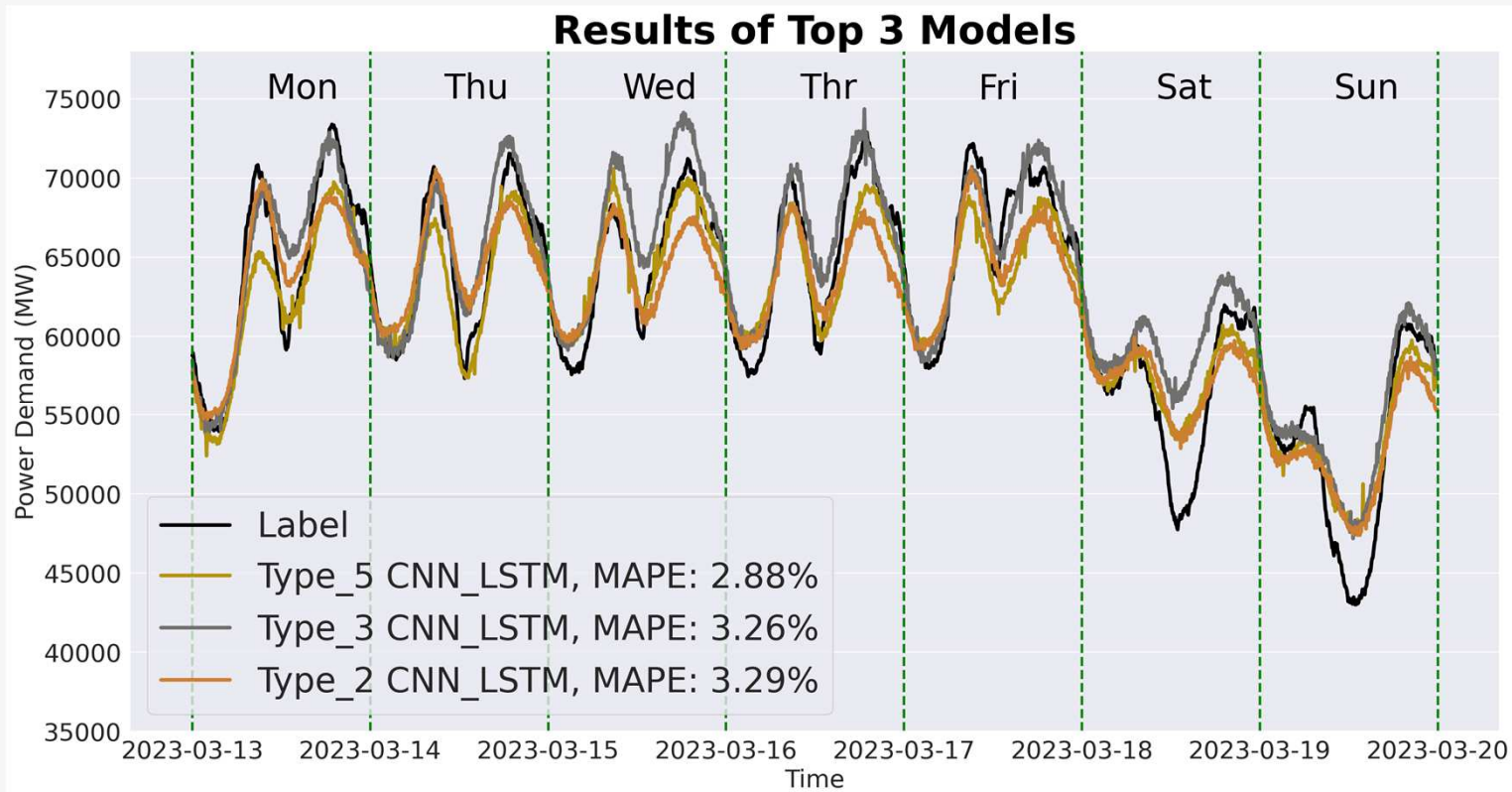
전체모델 MAPE (fig. 1)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

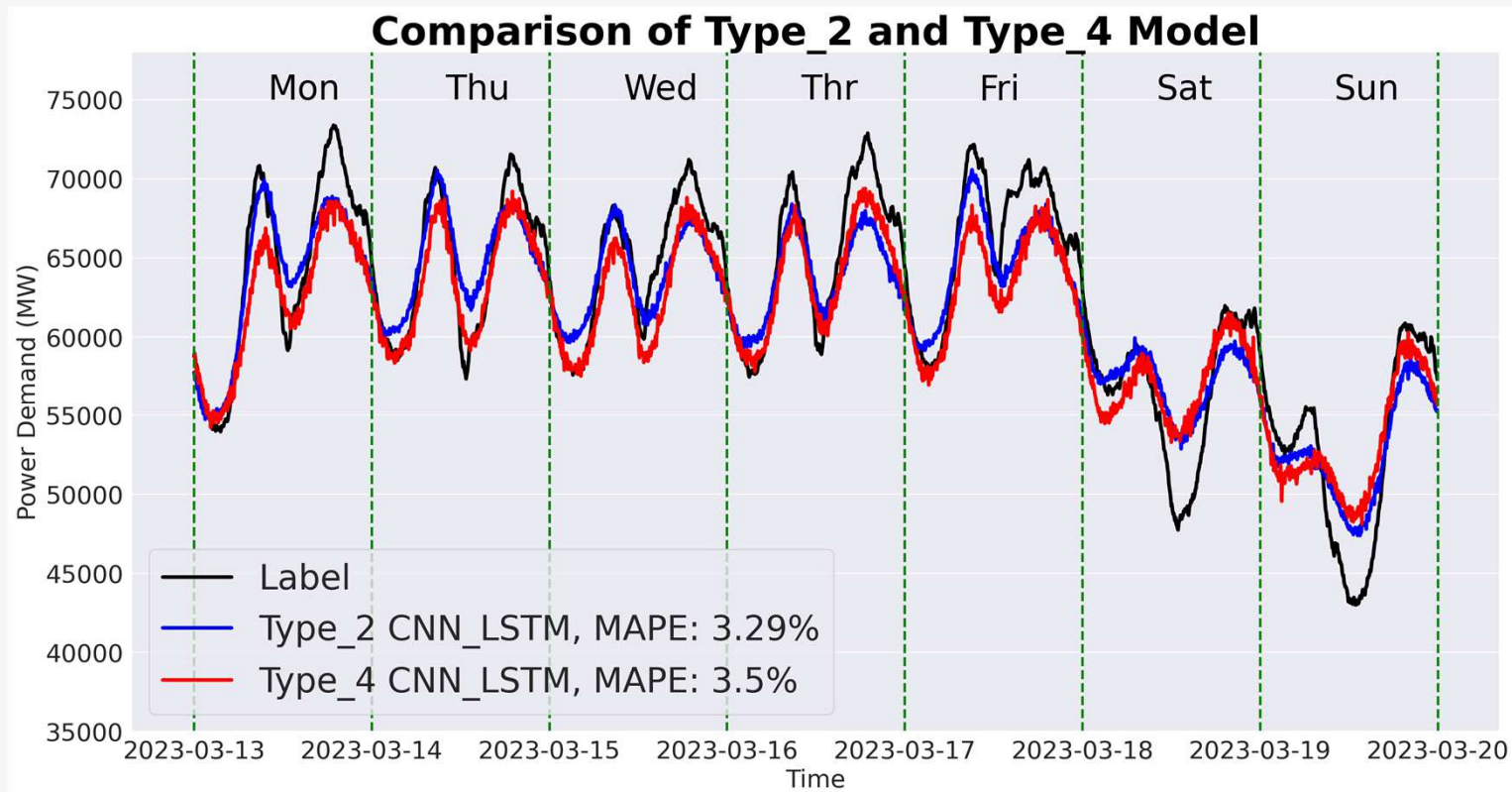
Top 3 Models MAPE (fig. 2)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

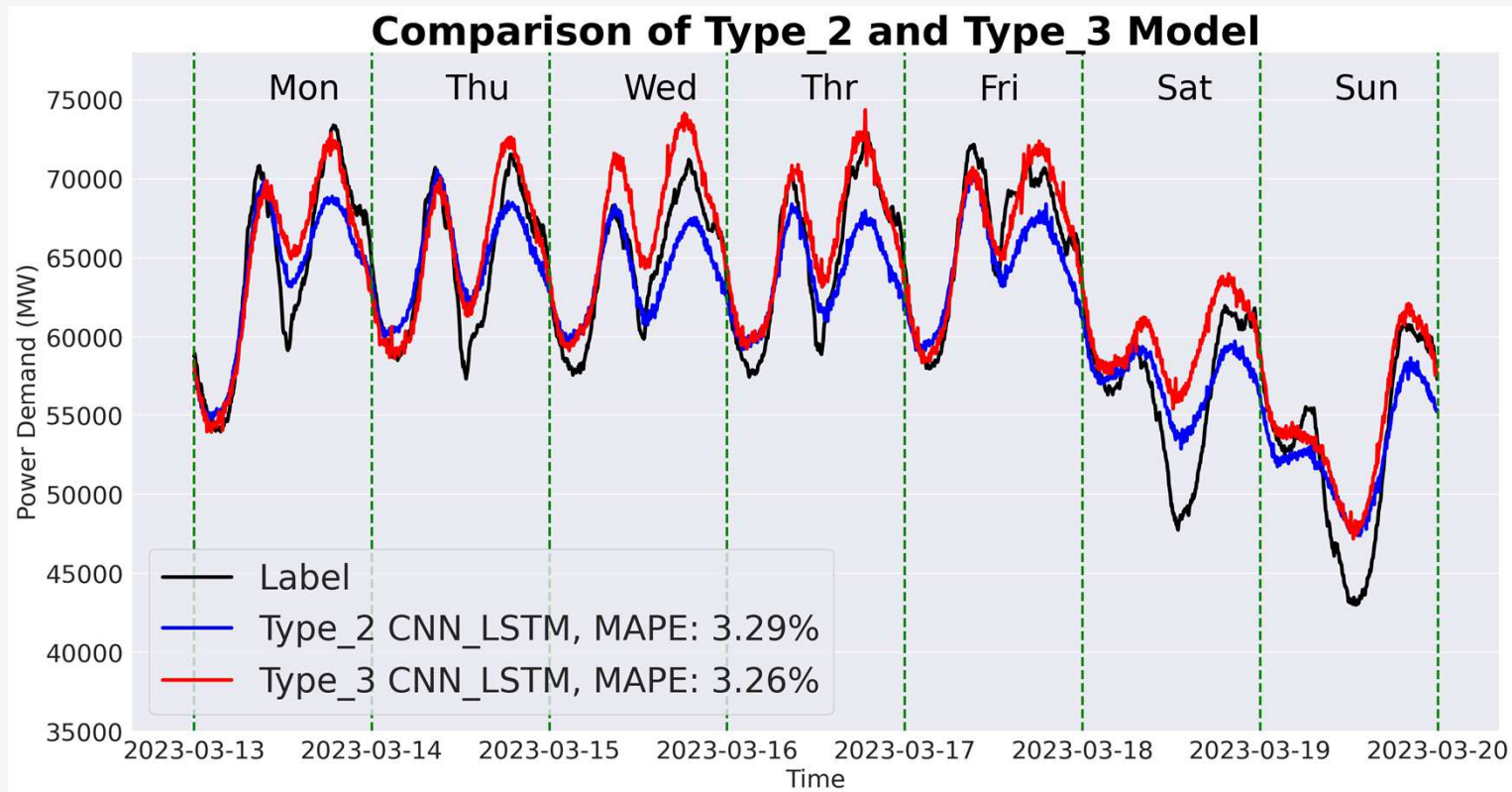
Type2 vs Type4 (fig. 3)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

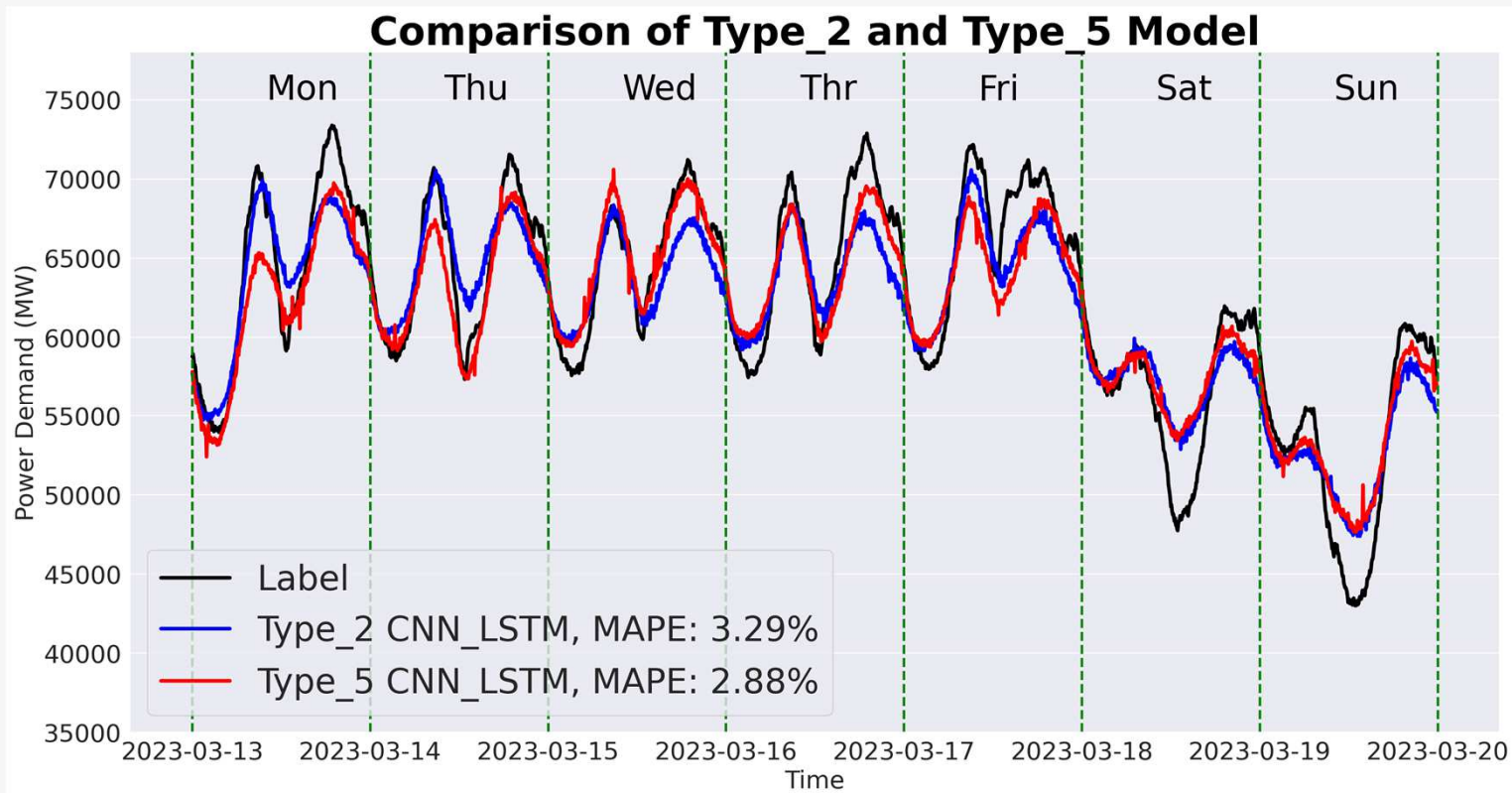
Type2 vs Type3 (fig. 4)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

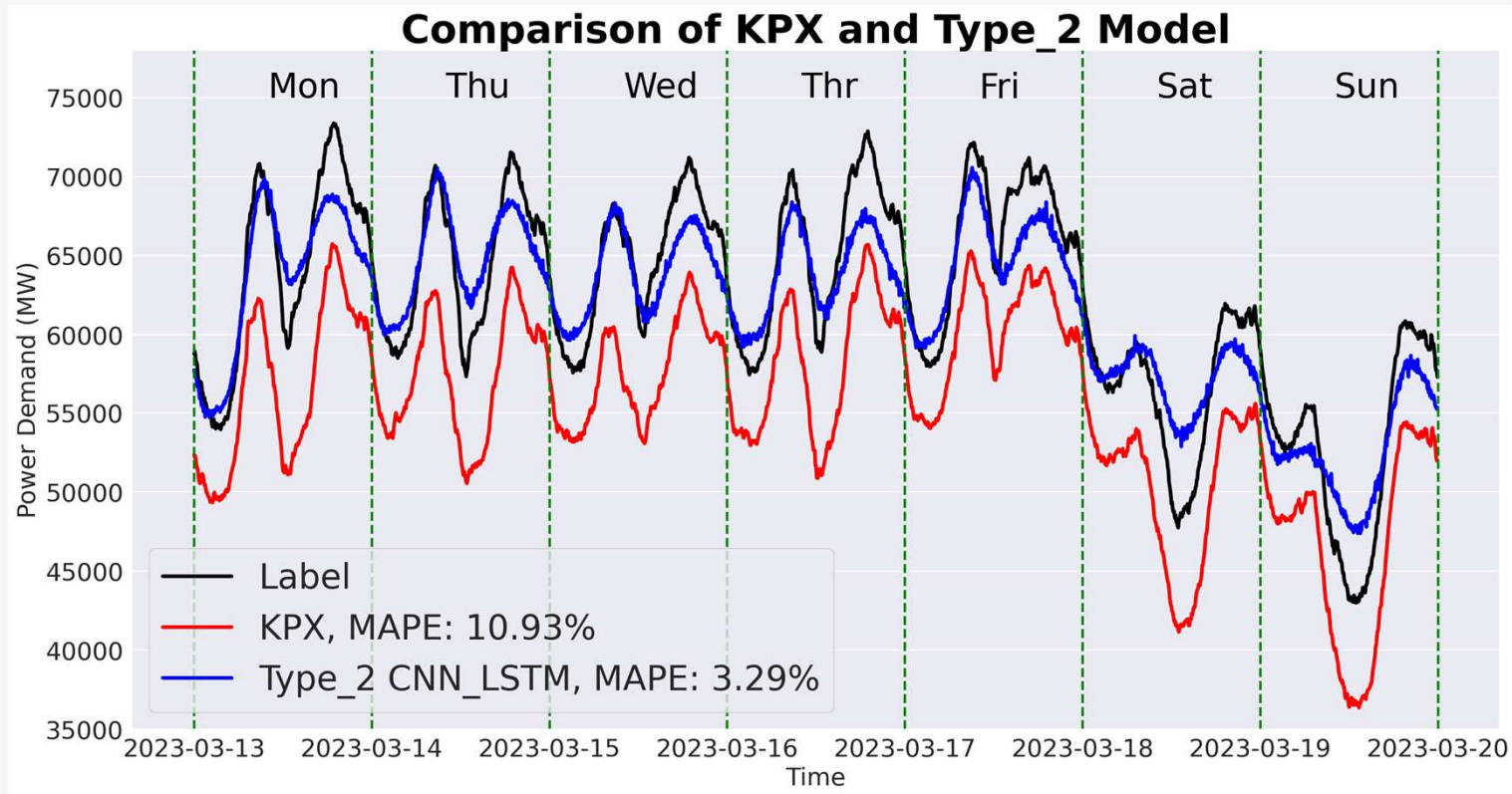
Type2 vs Type5 (fig. 5)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

시각 자료

Type2 vs KPX (fig. 6)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표

부록

최적의 모델 구성하기

최적의 모델 구성하기

요일별 최고 성능 모델을 조합하여 최적의 모델을 구성. Type3의 빈도가 가장 높다.

각 요 일별 최고성능 모델

- A. Mon
 - Type3 CNN_BiLSTM (2.748%)
- B. Thu
 - Type5 CNN_BiGRU (1.963%)
- C. Wed
 - Type5 CNN_BiLSTM (1.575%)
- D. Thr
 - Type3 CNN_BiLSTM (2.366%)
- E. Fri
 - Type3 CNN_BiLSTM (1.160%)
- F. Sat
 - Type4 CNN_BiLSTM (3.112%)
- G. Sun
 - Type5 CNN_BiLSTM (4.030%)

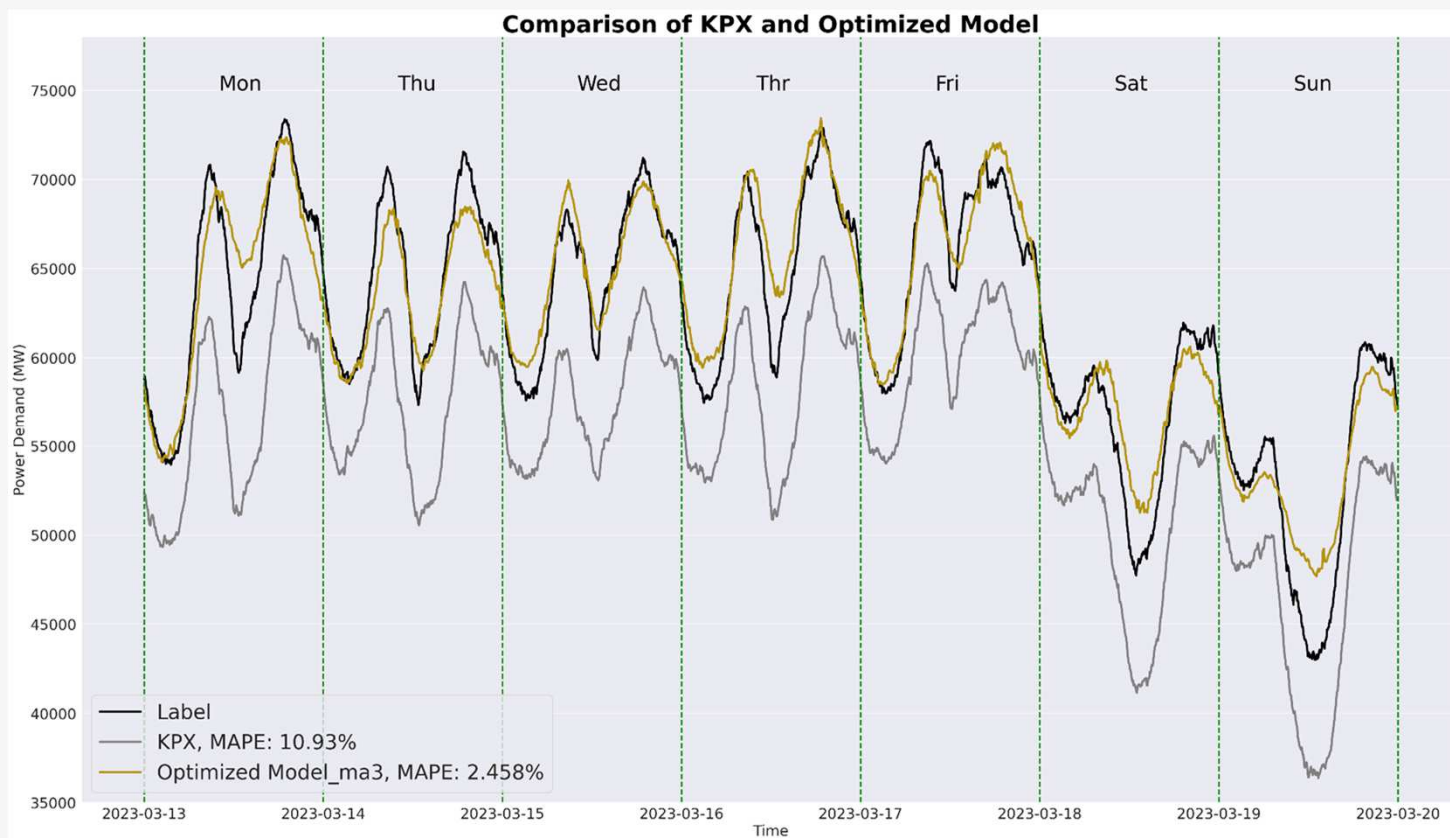
각 요 일별 최고성능 모델 선별

- Type3 : 3개 요일 (월, 목, 금)
- Type4 : 1개 요일 (토)
- Type5 : 3개 요일 (화, 수, 일)
- 모든 정보를 사용하는 Type5를 제외하면 기상정보를 사용하지 않는 Type3가 가장 많이 선택 되었다.

후처리

- 15분 이동평균선을 이용하여 노이즈의 영향을 줄인다.

Optimized Model vs KPX (fig. 7)



Type	Feature
# Type 1	전력 수요
# Type 2	전력수요, 달력
# Type 3	전력수요, 달력, 전력 보조지표
# Type 4	전력수요, 달력, 날씨
# Type 5	전력수요, 달력, 날씨, 전력 보조지표