



BDA 채용연계 데이터분석 공모전

제주도 신재생 에너지 활용을 통한 전력 비용 절감 및 탄소 배출 감소 분석

Team 브다히어로



목차 페이지

1

아이디어 개요

1. 문제 배경 및 필요성
2. 연구 목표

2

데이터 분석

1. 데이터 수집
2. 데이터 전처리
3. 분석

3

모델링

1. 전처리
2. 모델 소개
3. 초기 모델 결과
4. 개선 방안
5. 개선 후 성능 평가

4

기대효과 및 결론

1. 기대효과 및 결론

01 아이디어 개요

-1. 문제 배경 및 필요성

□ (탄소중립) 탄소중립도시 조성 및 환경자산의 현명한 이용

- 온실가스감축인지 예산제 확대, 기후변화영향평가 협의권한 제주특별법 이양
- 제주환경보전분담금 제도 입법화 추진(22대 국회의원 후보자 대상 입법안 제시)
- 꽃자왈 도민 자산화사업(13㏊, 20억 원) 및 송악산 사유지 매입(20㏊, 191억 원)

□ (그린수소) 그린수소로의 에너지대전환 기틀 마련

- 그린수소 생산 · 보급 · 활용 전주기 시설 확충
 - (수소생산) 그린수소 생산(3.3㎿, 12.5㎿, 30㎿), 수소 다변화(감귤박, 가축분뇨 등)
 - (보급활용) 민간 수소충전소 구축 지원(2개소), 수소차 보급 확대(41대)
- 발전 · 운송 · 생활 등 수소활용 확장여건 마련
 - 수소 혼전소 전환 명문화(제11차 전기본 반영 중), 상용차(버스 · 지게차 등)도입 등



- 최근 제주 지역에 태양광 설비와 원자력 발전소가 증가하면서 전력 생산이 급증. 그러나 전력 수요는 이 증가를 따라가지 못해 공급 과잉이 발생하는 상황.
- 이로 인해 대정전 위험이 증가하고 있으며, 효율적인 전력 관리가 필요.

- 제주도는 대한민국에서 신재생 에너지의 선도적인 역할을 수행하고 있으며, 2030년까지 CARBON FREE ISLAND를 목표로 정책을 추진 중.
- 제주도는 풍력과 태양광 같은 풍부한 자연 자원을 보유하고 있어 신재생 에너지 활용에 최적화된 지역.
- 신재생 에너지를 통해 화석 연료 의존도를 줄이고 탄소 배출을 줄이는 것은 전력 수요 공급 불일치 문제 해결에 긍정적 영향을 끼칠 것.

01 아이디어 개요

-1. 문제 배경 및 필요성

크롤링 및 감성 분석

- 네이버 뉴스기사 크롤링
 - 크롤링(CRAWLING)은 웹에서 자동으로 정보를 추출하는 과정으로, 데이터를 직접 수집하기 위해 활용한다. 본 연구에서는 제주도의 에너지 관련 인식 파악을 위한 감성 분석 데이터를 구하고자 HTML파일을 쉽게 파싱하고 추출을 돋는 BEAUTIFULSOUP으로 네이버 뉴스기사 크롤링을 진행했다.

1

검색 키워드 및 크롤링 기간 설정

ex) 검색 키워드 : 금산인삼축제크롤링
기간 : 2023.07.14 ~ 10.13 (축제 시작
전 3개월)

2

URL 크롤링 함수 정의

검색 결과 페이지 순차적 요청
네이버 뉴스 URL만 필터링

3

뉴스 기사 정보 크롤링

Beautifulsoup으로 html구조 파싱각
기사의 신문사, 제목 본문 크롤링

01 아이디어 개요

-1. 문제 배경 및 필요성

크롤링 및 감성 분석

- 네이버뉴스 제목 감성분석
 - 감성 분석(SENTIMENT ANALYSIS)은 텍스트 데이터를 분석하여 그 속에 담긴 감정이나 감정을 파악하는 자연어 처리(NLP)의 한 분야이다. 크롤링한 뉴스기사 제목은 핵심적인 정보를 담고 있으며, 해당 기사에서 전달하고자 하는 주요 감정을 반영한다. 본 연구에서는 제주도 에너지에 대한 긍정, 부정, 중립적인 감정을 뉴스기사 제목을 바탕으로 분석해, 제주 에너지에 대한 전반적인 여론과 인식을 파악했다.

1

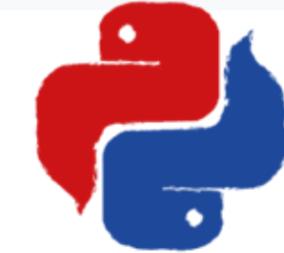
네이버 CLOVA
Sentiment API



뉴스 제목에 대한 감성분석 진행
감성 라벨, 신뢰도, 주요 감정 구간 추출

2

주요 감정 구간 TF-IDF
벡터화



KoNLPy Okt를 활용한 형태소 분
석제목별 최고 기여 단어 추출

3

데이터 정리

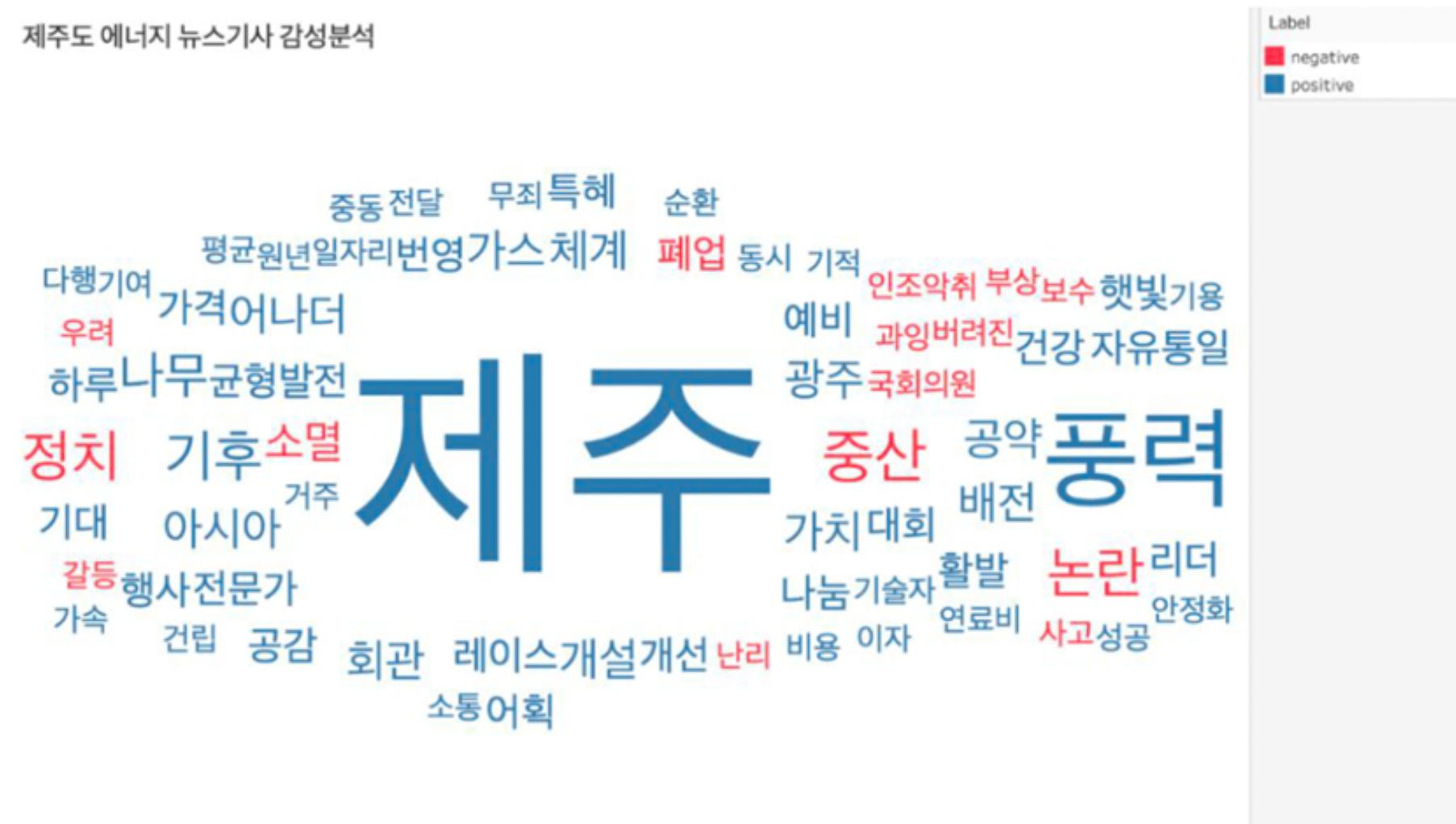
주요 감정 구간 형태소 단위 카운팅
중립 단어 제외한 긍부정 단어 필터링

01 아이디어 개요

-1. 문제 배경 및 필요성

크롤링 및 감성 분석

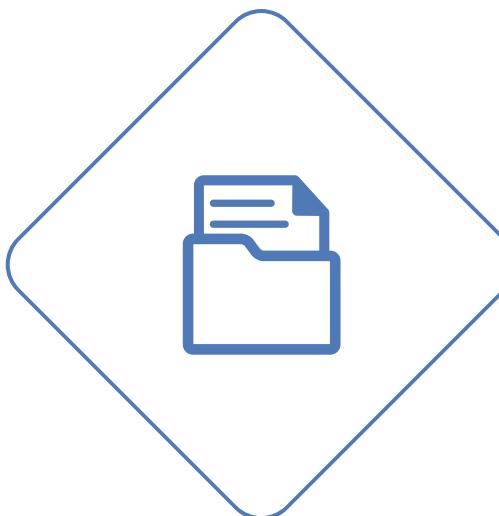
- 아래 '제주도 에너지'를 키워드로 검색한 네이버 뉴스기사 감성분석 결과 전반적인 제주도 에너지 문제에 대해 긍정적인 여론과 인식이 보여진다. 정부는 신재생 에너지 프로젝트에 대한 정책적, 재정적 지원을 제공하고 있고, 지역 사회의 인식 변화와 참여가 활성화되고 있어, 본 연구가 설정한 주제 또한 이러한 참여에 실질적인 기여가 될 것이다.



[TABLEAU를 활용한 워드클라우드 시각화]

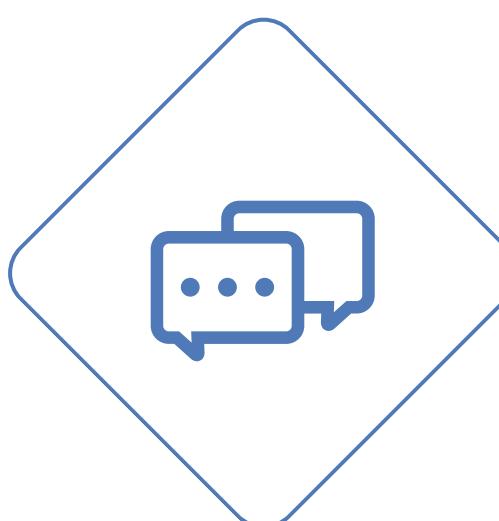
01 아이디어 개요

-2. 연구 목표



예측 모델 개발

기상 데이터와 태양광, 풍력 발전량의 30일의 데이터를 입력으로 사용하여 다음 1일의 전력 공급량을 단기 예측하는 시계열 모델을 개발한다.



효과 분석

예측 모델과 연료원별 정산단가 데이터를 통해 소비자와 공급자 모두에게 전력 비용 절감 효과와 탄소 배출 감소 효과를 분석한다.

02 데이터 분석

-1. 데이터 수집

구분	제공기관	데이터명	내용
1	전력데이터 개방 포털시스템	제주시 가구 평균 전력 사용량	지역구분, 대상가 구수(호), 가구당 평균 사용량(kWh), 가구당 평균 전기요금(원)
2	공공데이터포털	제주에너지공사 _ 풍력발전기 전력 정보	일시, 가시리풍력, 김녕풍력, 동북풍력, 풍력전체 등
3	공공데이터포털	제주에너지공사 _ 태양광발전기 전력 정보	일시, 교래태양광, 종합경기장 태양광, 태양광전체 등
4	기상자료개방포털	조건별 기후통계	일시, 평균기온, 강수량, 풍속, 습도 등
5	전력통계 정보시스템	연료원별 정산단가	원자력, 유류, 신재생 에너지 등 별 정산단가

02 데이터 분석

-2. 데이터 전처리

1. 제주시 전력 사용량



가구당 평균 사용량, 전기요금 등

제주 주민들의 평균 전력 수요를 파악하고자
날짜별 데이터 수집

3. 제주 기상 통계



제주 날씨 정보

날짜별 제주 평균기온, 풍속, 습도 등
기상 정보 데이터 수집

5. 최종 데이터셋

독립 변수

모두 날짜별로 수집한 가구당 평균 전력
사용량, 기상 정보

종속 변수

날짜별 태양열 발전량, 풍력 발전량

2. 풍력, 태양열 발전 정보



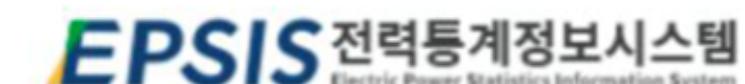
풍력 발전량

날짜별 풍력 발전량 데이터 수집

태양열 발전량

날짜별 태양열 발전량 데이터 수집

4. 연료원별 정산단가



연료원별 정산단가

원자력, 유류, 신재생 에너지 등 연료원별
정산단가 데이터 수집



추후 수요 공급 최적화 예측값에 따라 절약한
값을 계산하기 위함.

02 데이터 분석

-3. 분석

<상관 분석>

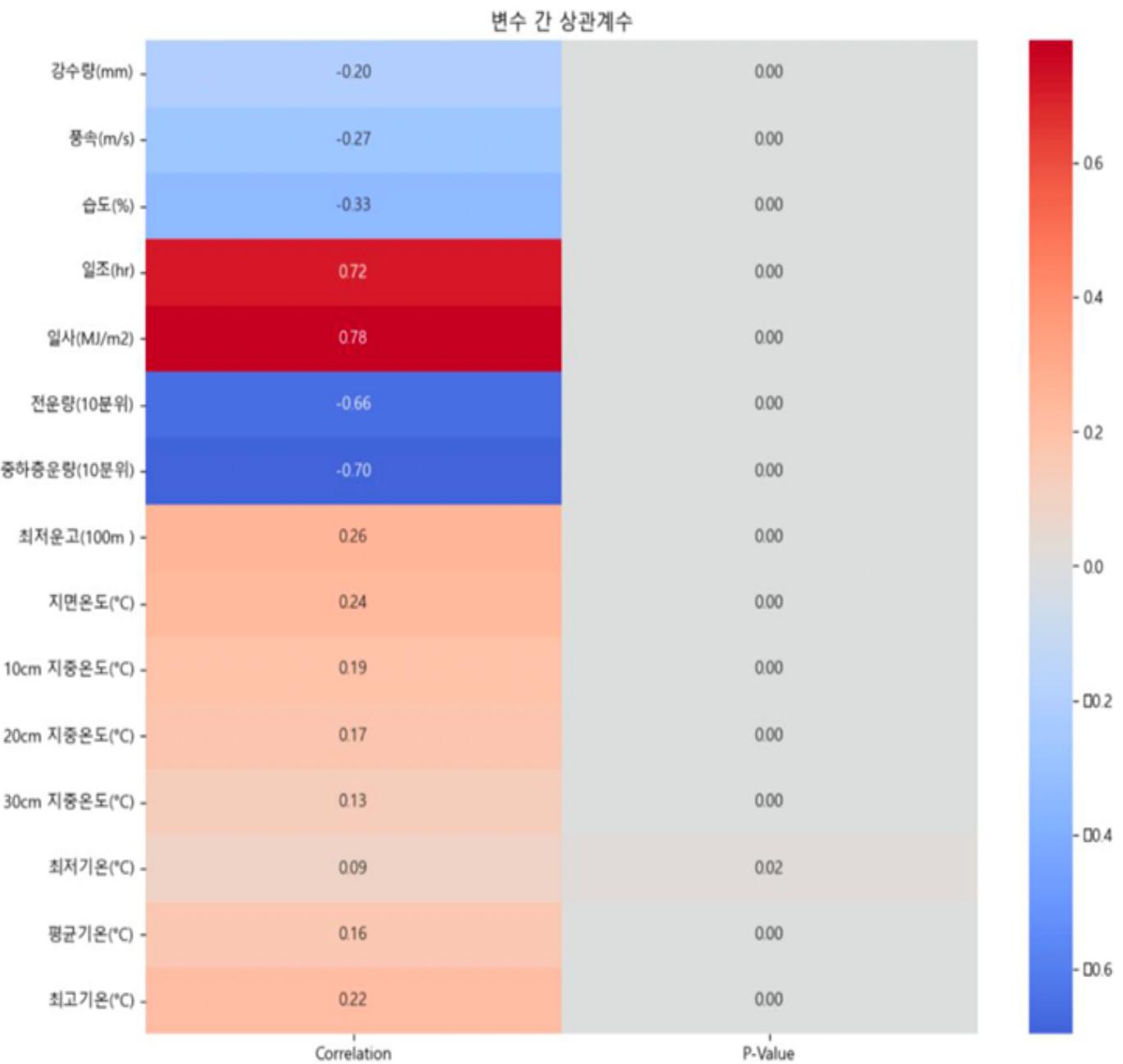
- 본 연구는 2022~2023년의 날짜별 총 730개 행 전처리 데이터를 바탕으로 발전량(=공급량)을 예측하는 모델링을 진행하기 위해 독립변수 - 종속 변수 간 상관관계 분석을 우선적으로 진행하여, 중요 피처만 선정하는 과정을 거쳤다.

1

기상데이터-태양광발전량 간의 상관관계

- Scipy 라이브러리의 피어슨 상관분석 메서드를 호출하여 상관분석을 진행한 결과는 다음과 같다.
- 다음과 같은 결과에서 상관계수 절대값이 0.5 이상이면서 p-value가 유의미한 독립변수만 채택하였다.
- 그 결과 **일조, 일사**가 태양광 발전량과 깊은 양의 상관관계가 있음을 확인하고 모델링에 포함될 독립변수로 선정하였다.

[상관분석 히트맵]



02 데이터 분석

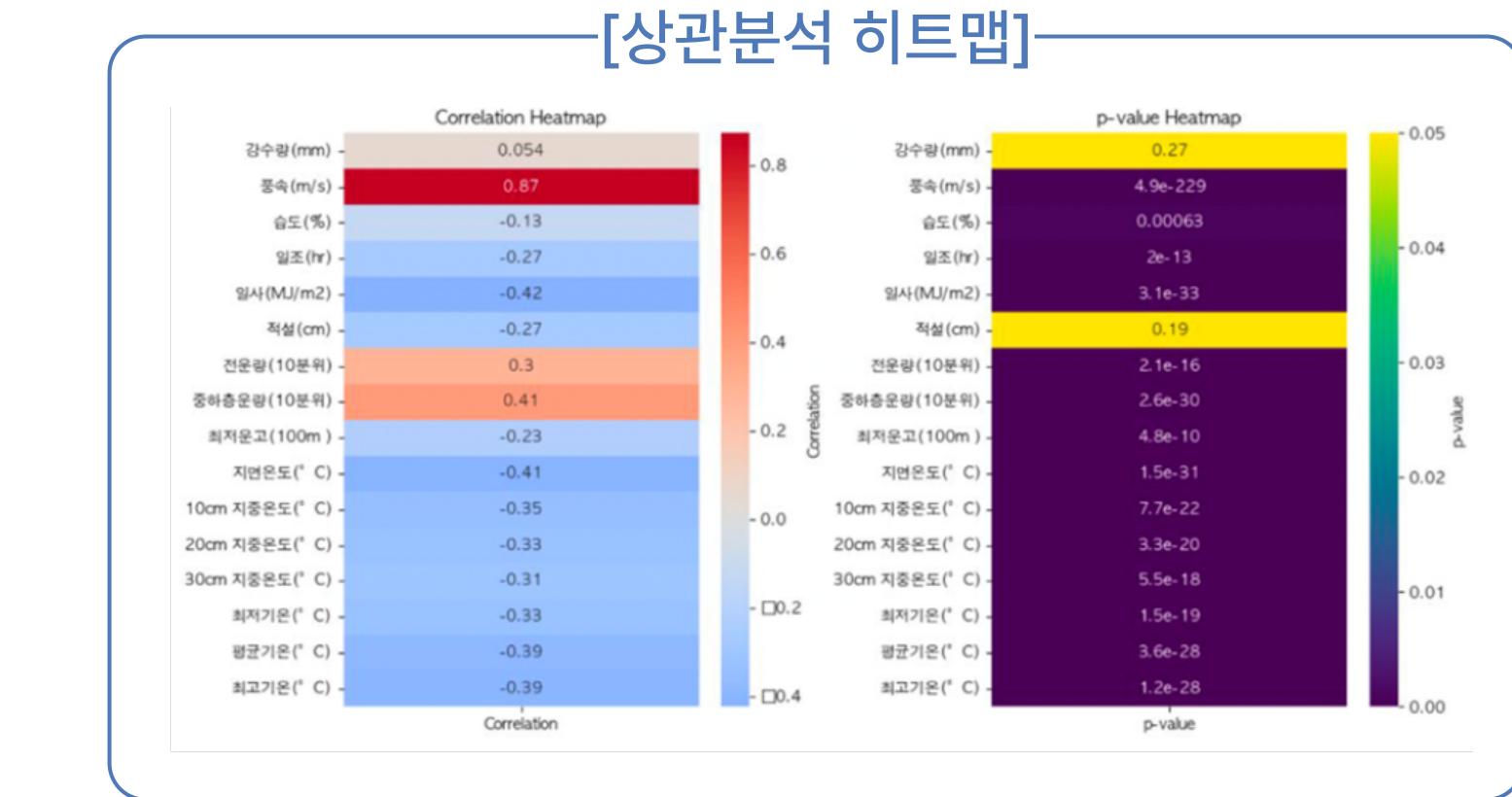
-3. 분석

<상관분석>

2

기상데이터-풍력발전량 간의 상관관계

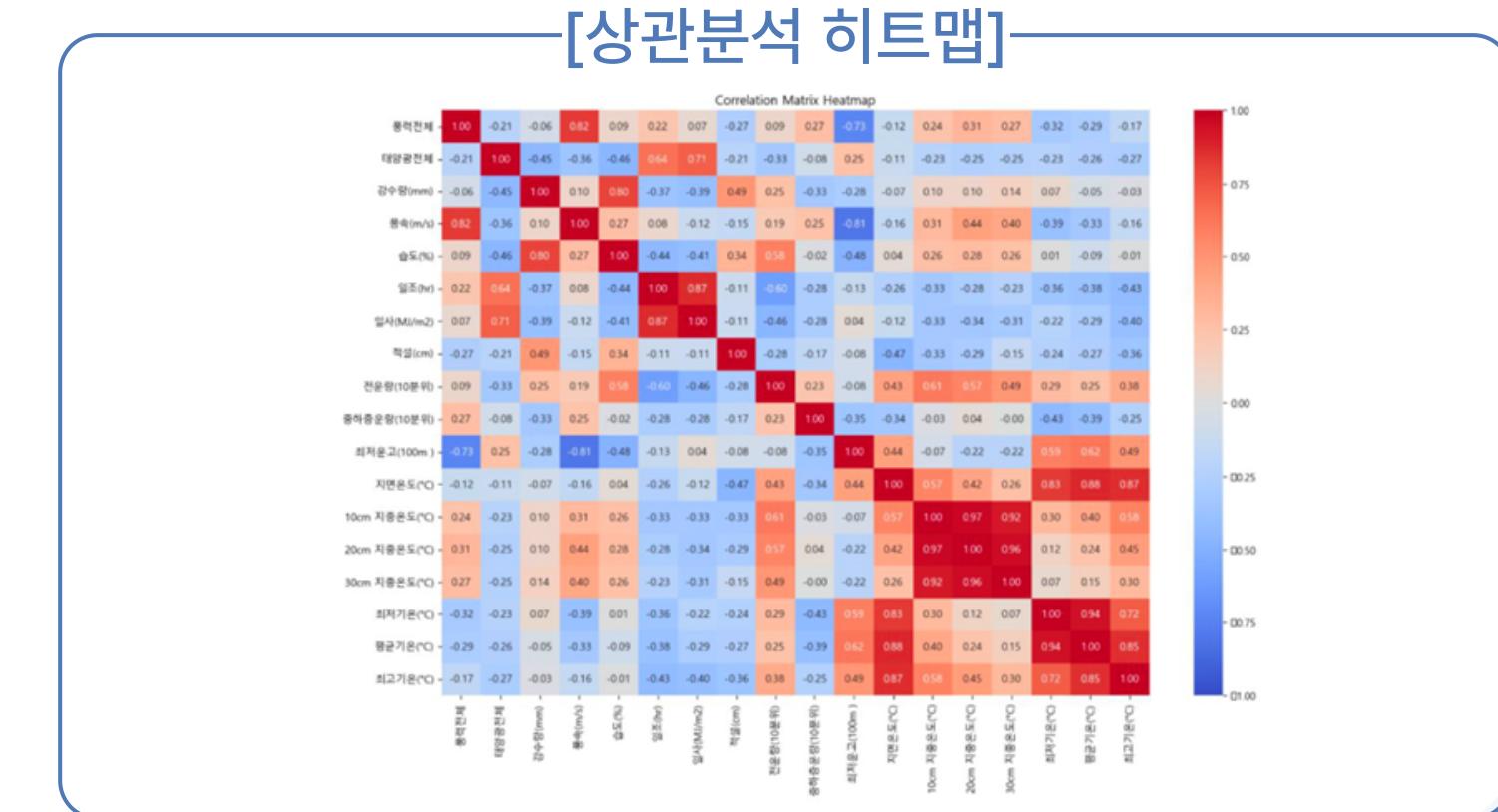
- 동일한 방법으로 기상데이터와 풍력발전량 간의 상관관계도 계산하였다. 그 결과 **풍속, 일사, 중하층운량, 지면온도**가 풍력 발전량과 깊은 상관관계가 있음을 확인하고 모델링에 포함될 독립변수로 선정하였다.



3

기상데이터-태양광+풍력발전량 병합 데이터 간의 상관관계

- 동일한 방법으로 기상데이터와 풍력발전량 간의 상관관계도 계산하였다. 그 결과 태양광에서 유의미하고 강한 상관관계 나타내는 기상 컬럼은 **일조, 일사**, 풍력에서 유의미하고 강한 상관관계 나타내는 기상 컬럼은 **풍속, 최저운고**임을 확인하고 모델링에 포함될 독립변수로 선정하였다.



03 모델링

-1. 전처리

데이터 전처리

[Feature Selection]

태양광

일조, 일사

풍력

풍속, 일사,
지면온도,
중하층운량

[Data Processing]

- 데이터 정제 :
결측치 0으로 대체
- 데이터 변환 :
일시 컬럼을 날짜 형식으로 변환
Min-Max 스케일링
LSTM/GRU 모델의 입력 형식에 맞게 시계열
데이터로 변환
- 데이터 분할 :
학습 및 테스트 데이터 셋으로 분할
- 조기 종료 콜백 설정 :
과적합 장비를 위한 조기 종료 콜백 설정

03 모델링

-2. 모델 소개

사용한 모델 종류

- 서로 다른 특성을 가진 모델들을 비교함으로써, 태양광과 풍력 발전량 예측이라는 복잡한 문제에 대한 접근 방식을 찾고자 5가지 모델을 선정

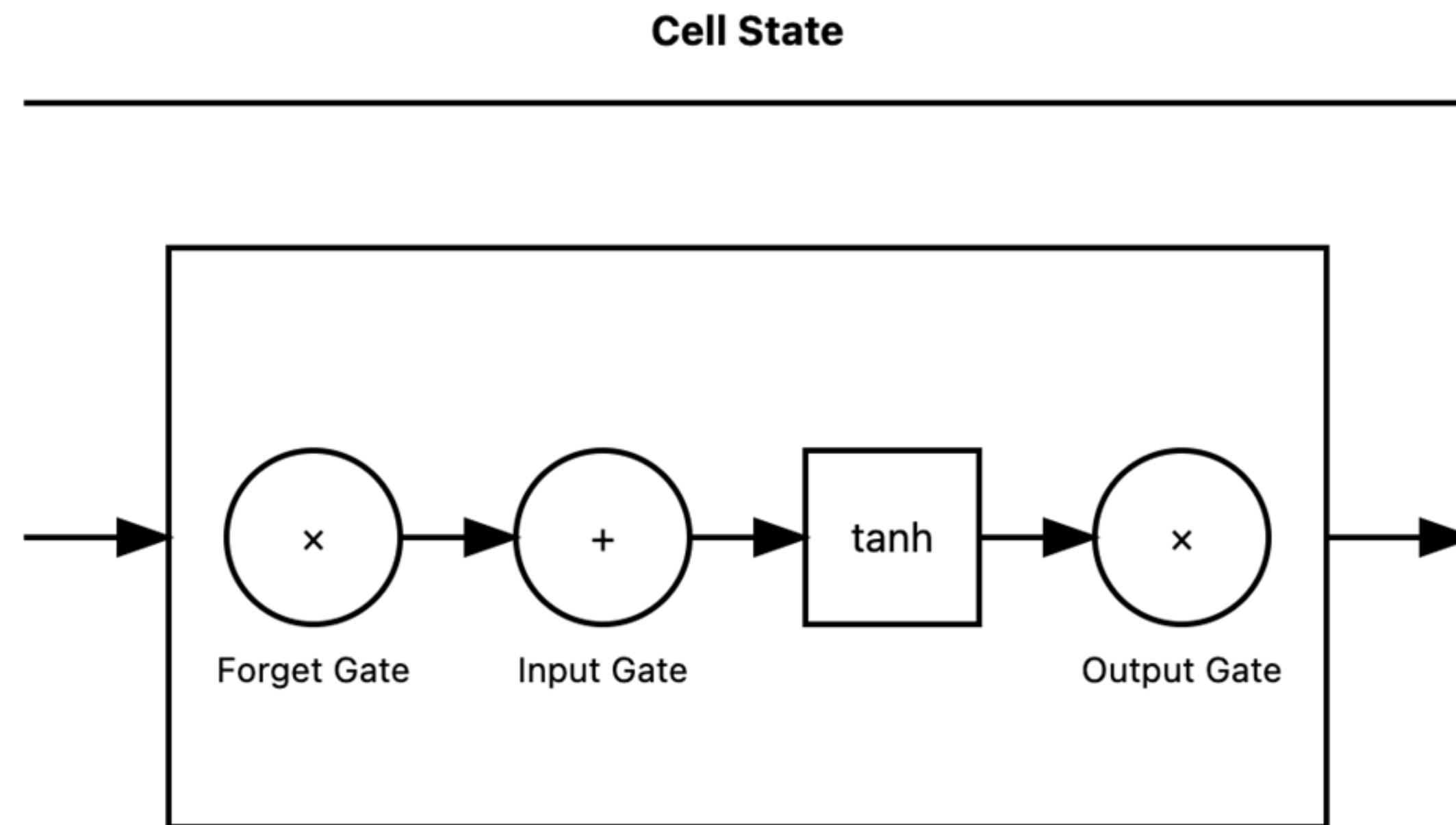
모델명	유형	주요 특징	용도
LSTM	딥러닝	장기 의존성 학습 가능	시계열 예측, 순차 데이터 처리
GRU	딥러닝	LSTM보다 단순한 구조	시계열 예측, 자연어 처리
Random Forest	앙상블	여러 결정 트리의 조합	분류, 회귀, 특성 중요도 평가
XGBoost	앙상블	그래디언트 부스팅 기반	분류, 회귀, 순위 예측
ARIMA	통계	자기회귀와 이동 평균의 결합	시계열 예측, 추세 및 계절성 분석

03 모델링

-2. 모델 소개

LSTM (Long-Short-Term Memory)

- '긴 단기 기억' 이라는 뜻으로 시계열 데이터의 장기 의존성을 학습하고 예측하는 신경망 모델

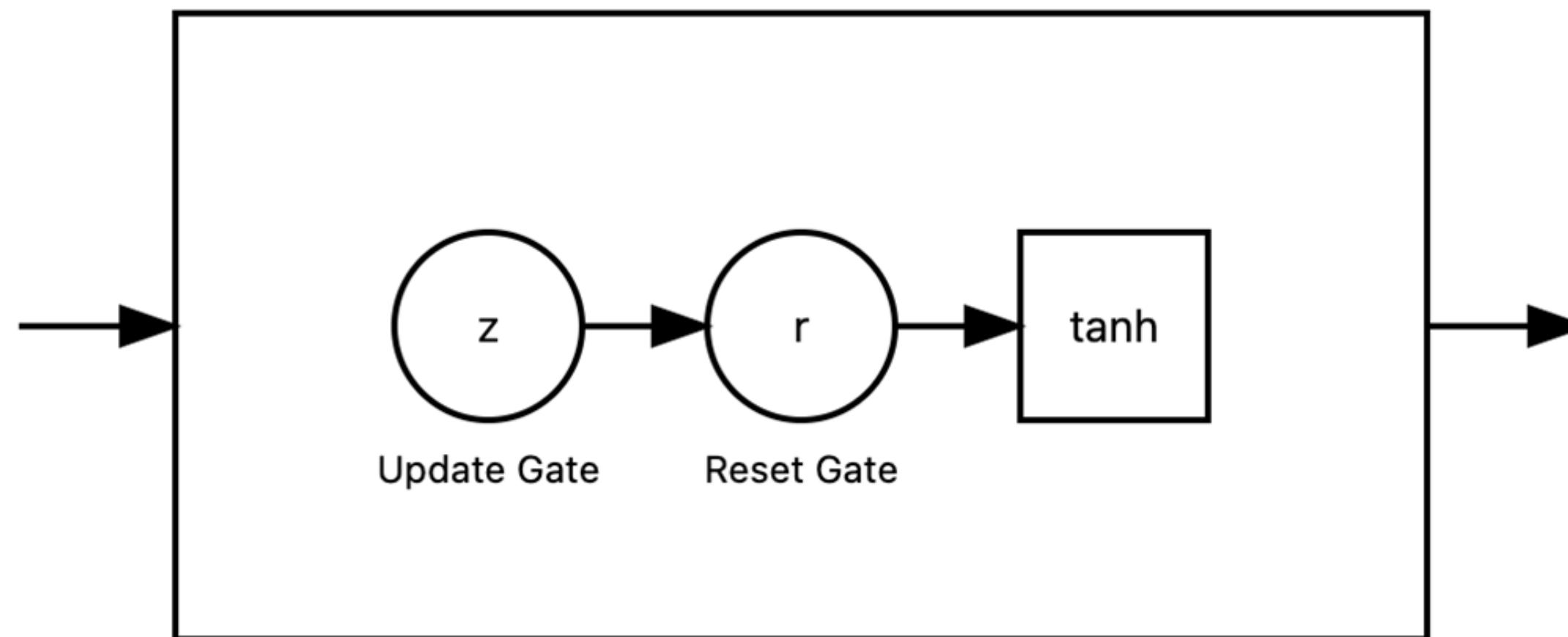


03 모델링

-2. 모델 소개

GRU (Gated Recurrent Unit)

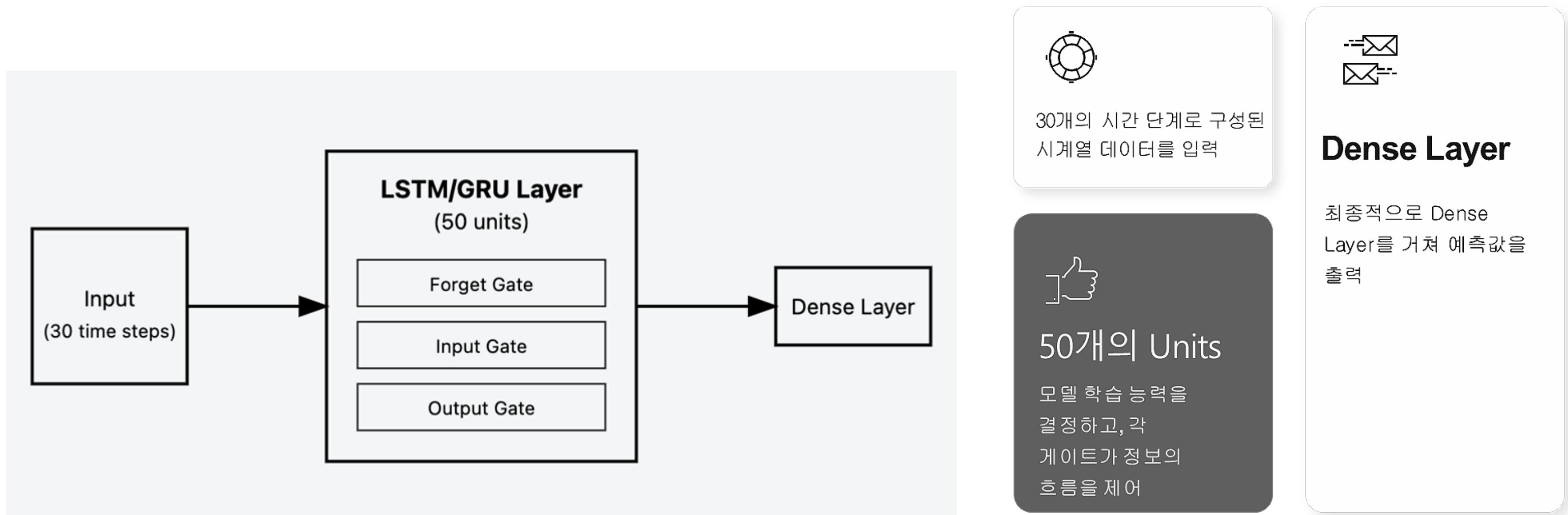
- LSTM보다 단순한 구조로, 계산 효율성이 높고
작은 양의 데이터로도 좋은 성능을 보이는 신경망
모델



03 모델링

-2. 모델 소개

LSTM과 GRU의 적용

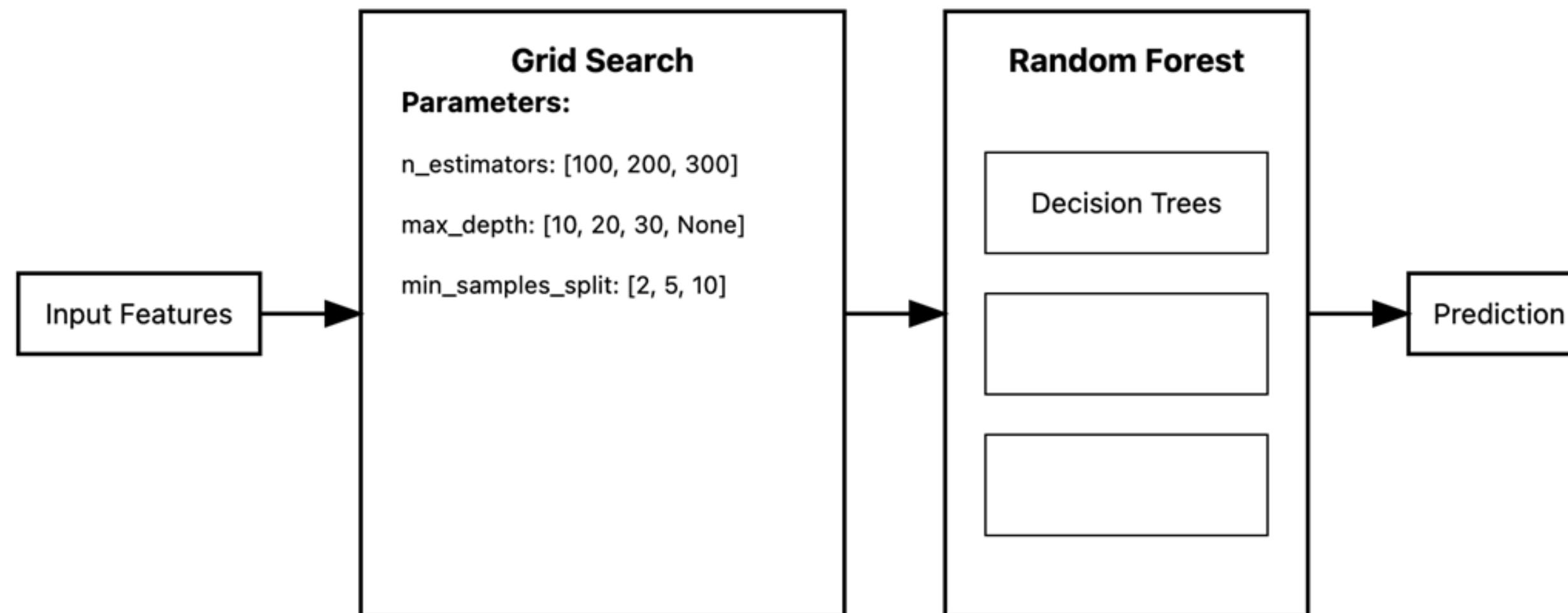


03 모델링

-2. 모델 소개

Random Forest

- 다수의 결정 트리를 생성하고 그 결과를 평균내어 예측을 수행하는 양상블 학습 모델

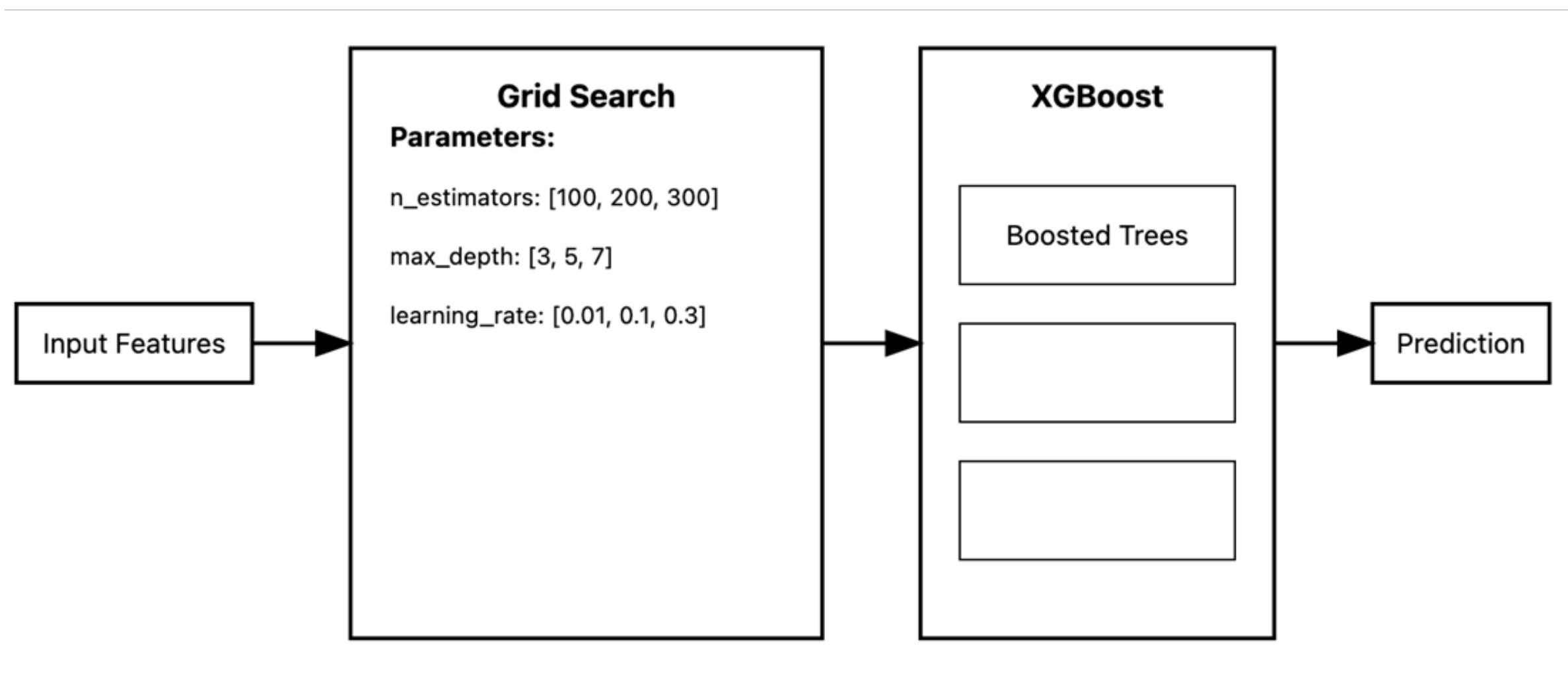


03 모델링

-2. 모델 소개

XGBoost (eXtreme Gradient Boosting)

- 그래디언트 부스팅 기반의 의사결정 트리 앙상블 모델로, 높은 예측 성능과 빠른 학습 속도

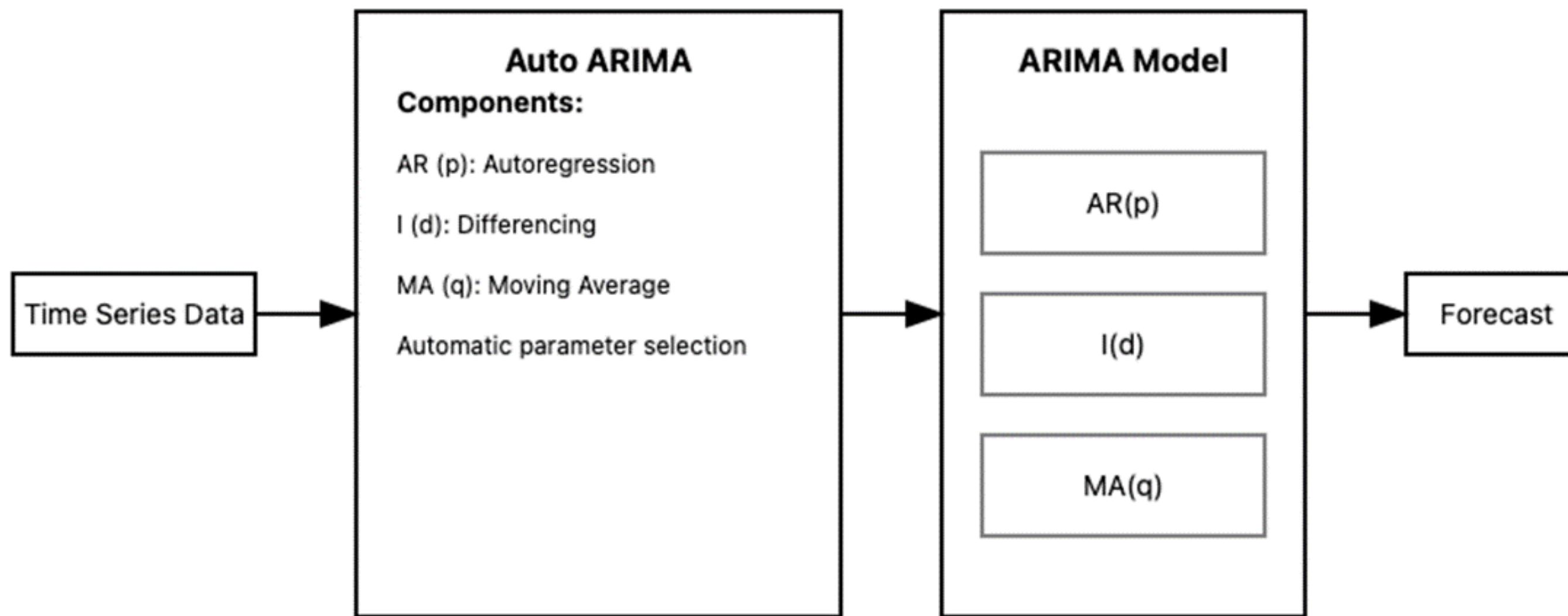


03 모델링

-2. 모델 소개

ARIMA (Autoregressive Integrated Moving Average)

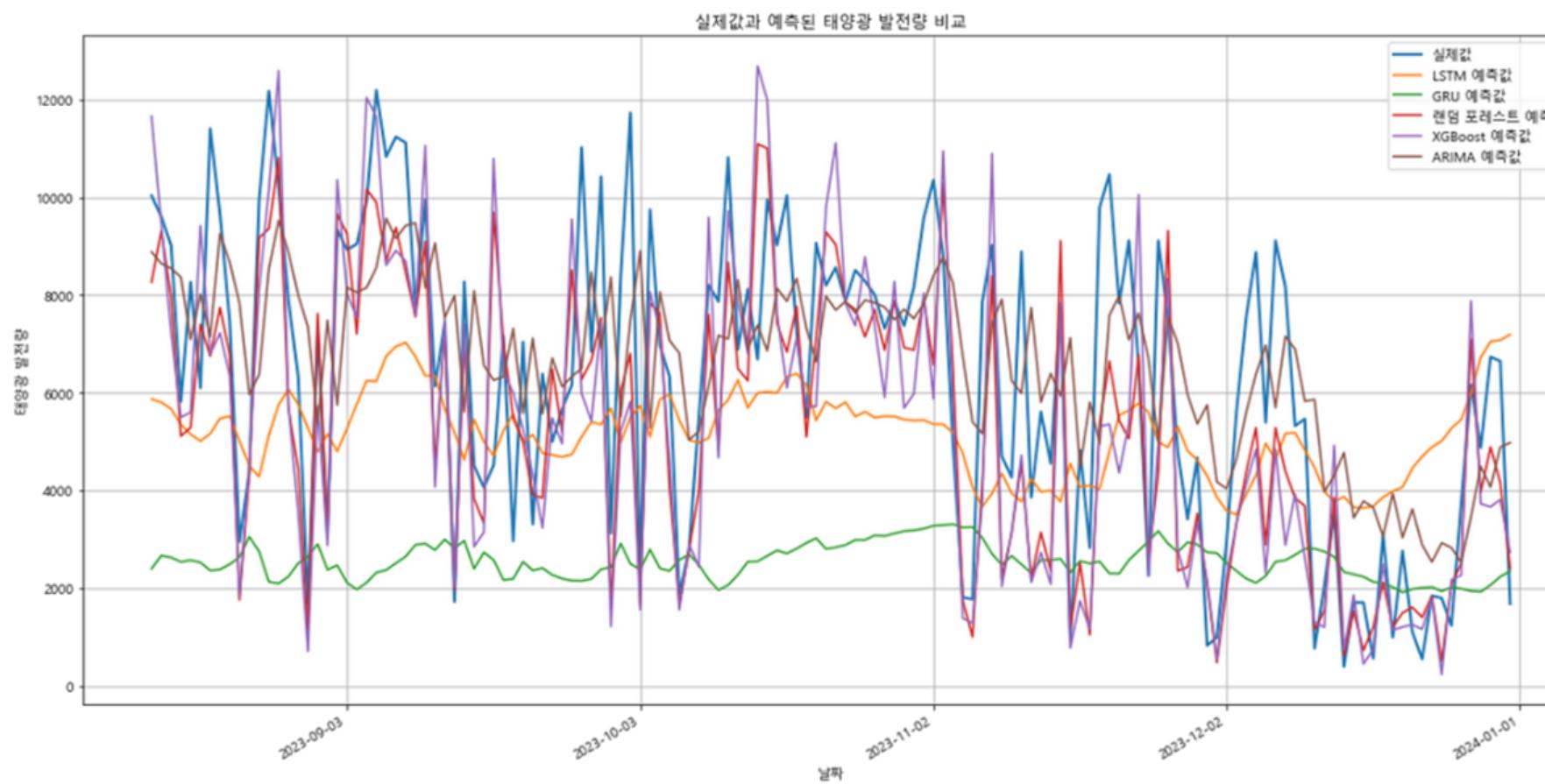
- 시계열 데이터의 자기회귀, 차분, 이동평균을 결합한 통계 모델로, 시간에 따른 추세와 계절성을 예측



03 모델링

-3. 초기 모델 결과

태양광 발전량 성능 평가



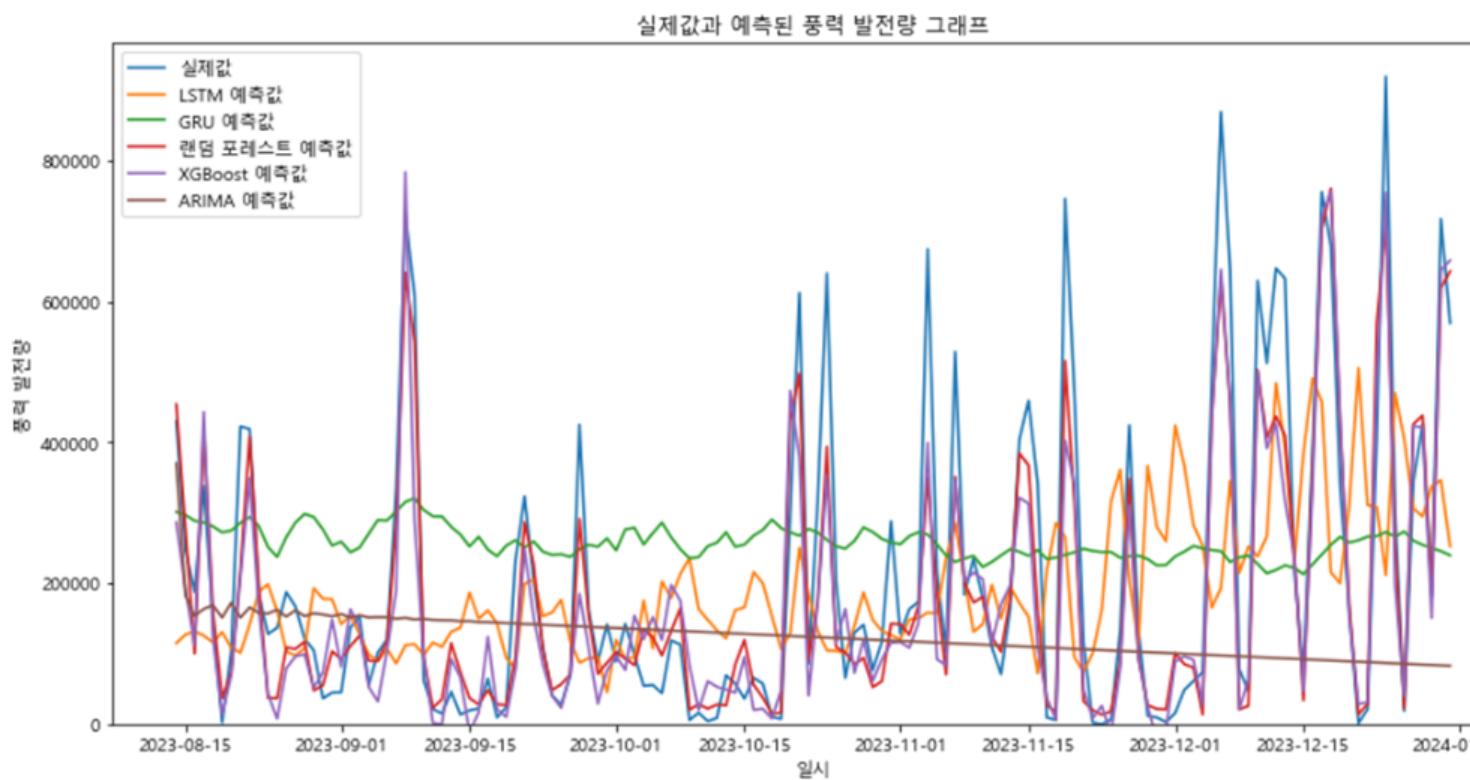
	MSE	MAE	NMAE	R2	Accuracy
LSTM	9.515340e+06	2608.674952	108.676677	0.012879	1.287902
GRU	2.358802e+07	4098.961306	170.761594	-1.447020	0.000000
Random Forest	3.909756e+06	1526.090972	63.576528	0.594402	59.440207
XGBoost	5.356240e+06	1827.474254	76.132072	0.444344	44.434384
ARIMA	6.928493e+06	2136.43181	89.003158	0.281238	28.123837

Random Forest 모델이 가장 우수한 성능을 보인다.

03 모델링

-3. 초기 모델 결과

풍력 발전량 성능 시각화



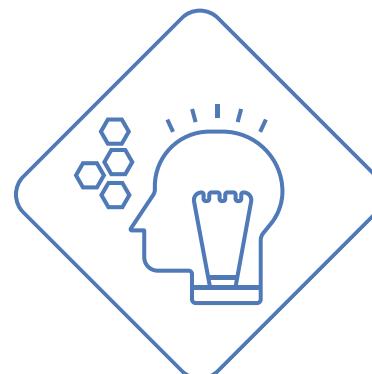
	MSE	MAE	NMAE	R2	Accuracy
LSTM	47473993574.148	159813.272770	292.275843091	0.1761323	17.613227
GRU	62653826224.760	210860.505959	377.485126609	-0.0872998	0.000000
Random Forest	7003426231.761	59557.721282	110.323505000	0.8680126	86.801263
XGBoost	9866394220.134	71706.938069	132.831872085	0.8140568	81.405681
ARIMA	70326094859.472	176844.297208	323.940395619	-0.3253736	0.000000

Random Forest 모델이 가장 우수한 성능을 보인다.

03 모델링

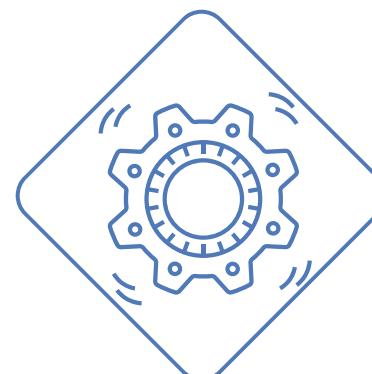
-4. 개선 방안

교차 검증과 그리드 서치



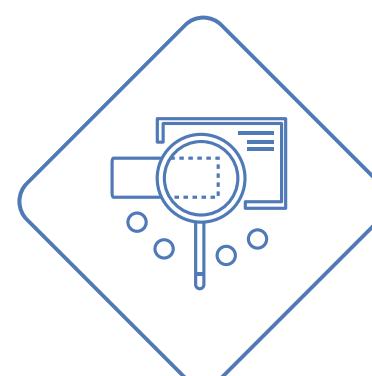
모델 개선 방안

다음 과정을 통해 단순히 모델 적용으로 그치는게 아닌 더 신뢰성이 있고, 성능이 우수하며, 실제 문제에 더 잘 적용할 수 있는 모델을 개발할 수 있다.



교차 검증

다데이터를 여러 부분으로 나누어 반복적으로 학습과 검증을 수행함으로써 모델의 일반화 성능을 평가한다.



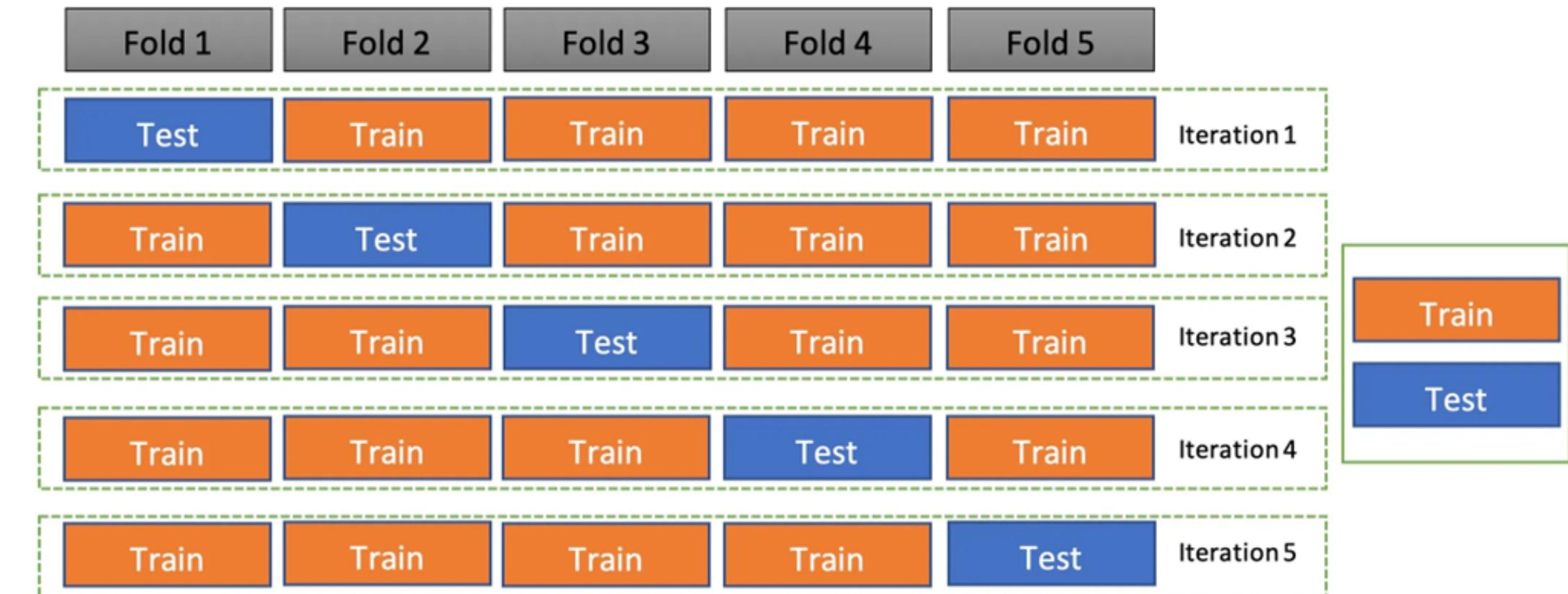
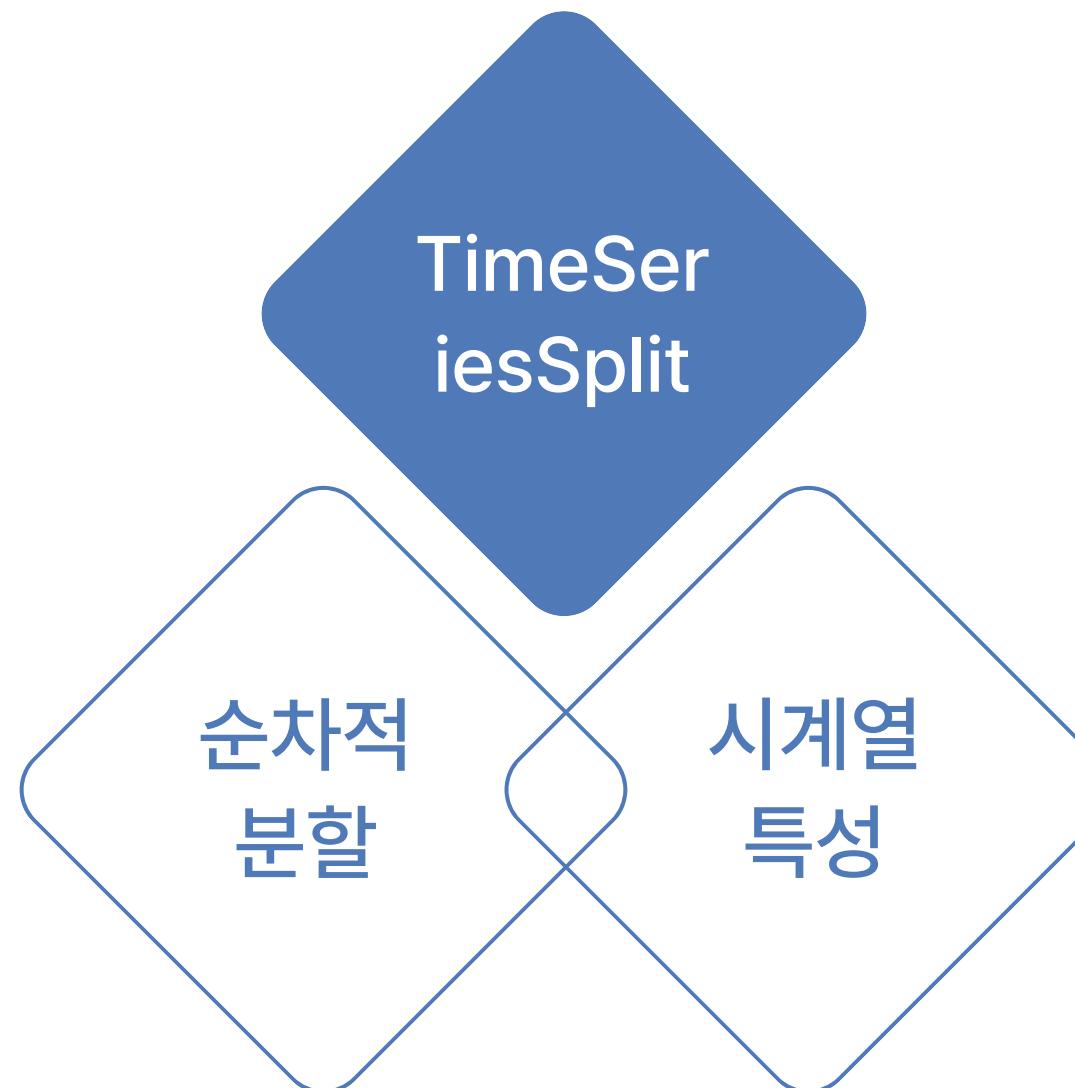
그리드 서치

다모델의 여러 하이퍼파라미터 조합을 체계적으로 시도하여 최적의 성능을 내는 조합을 찾는다.

03 모델링

-4. 개선 방안

교차 검증

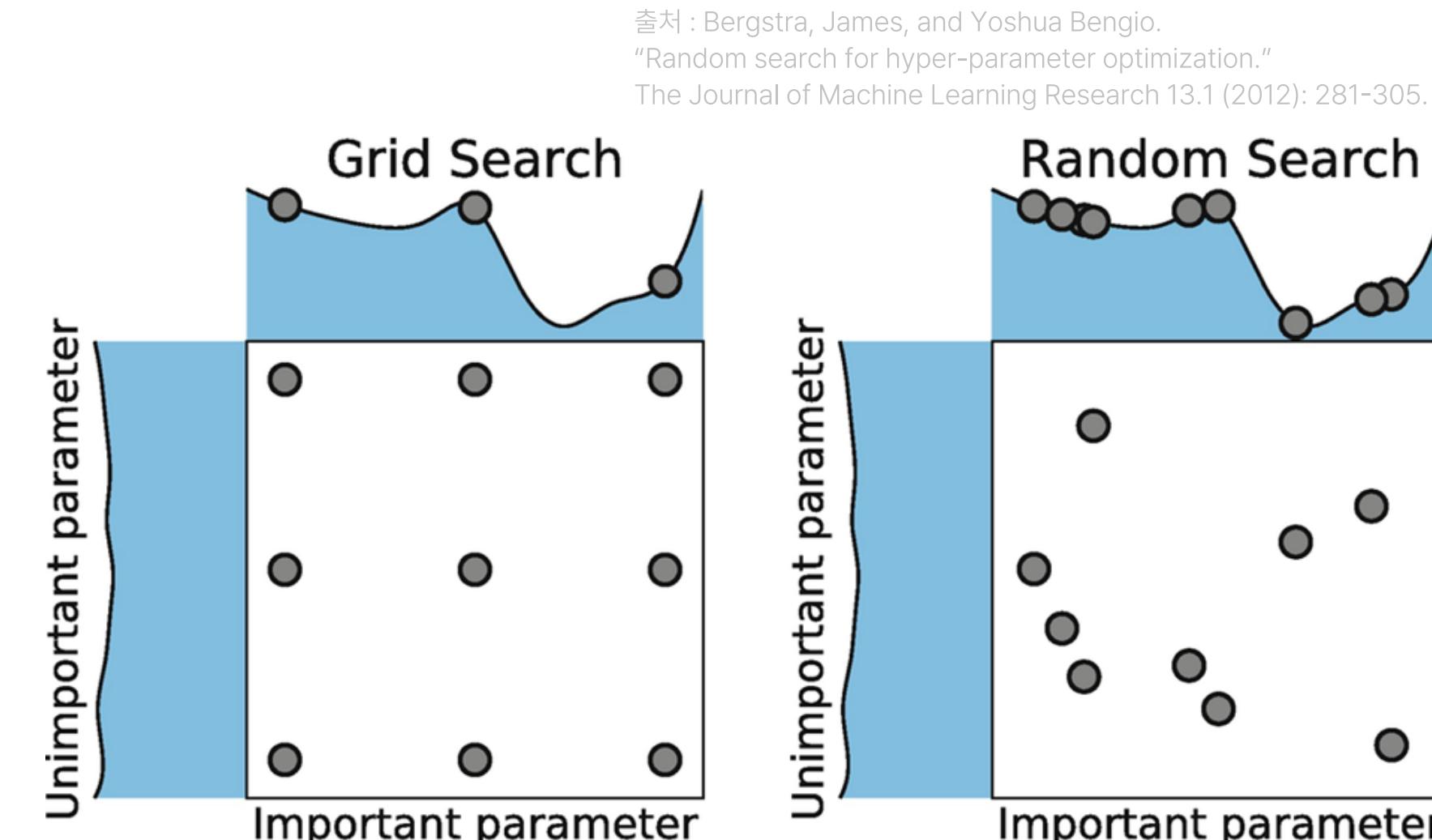
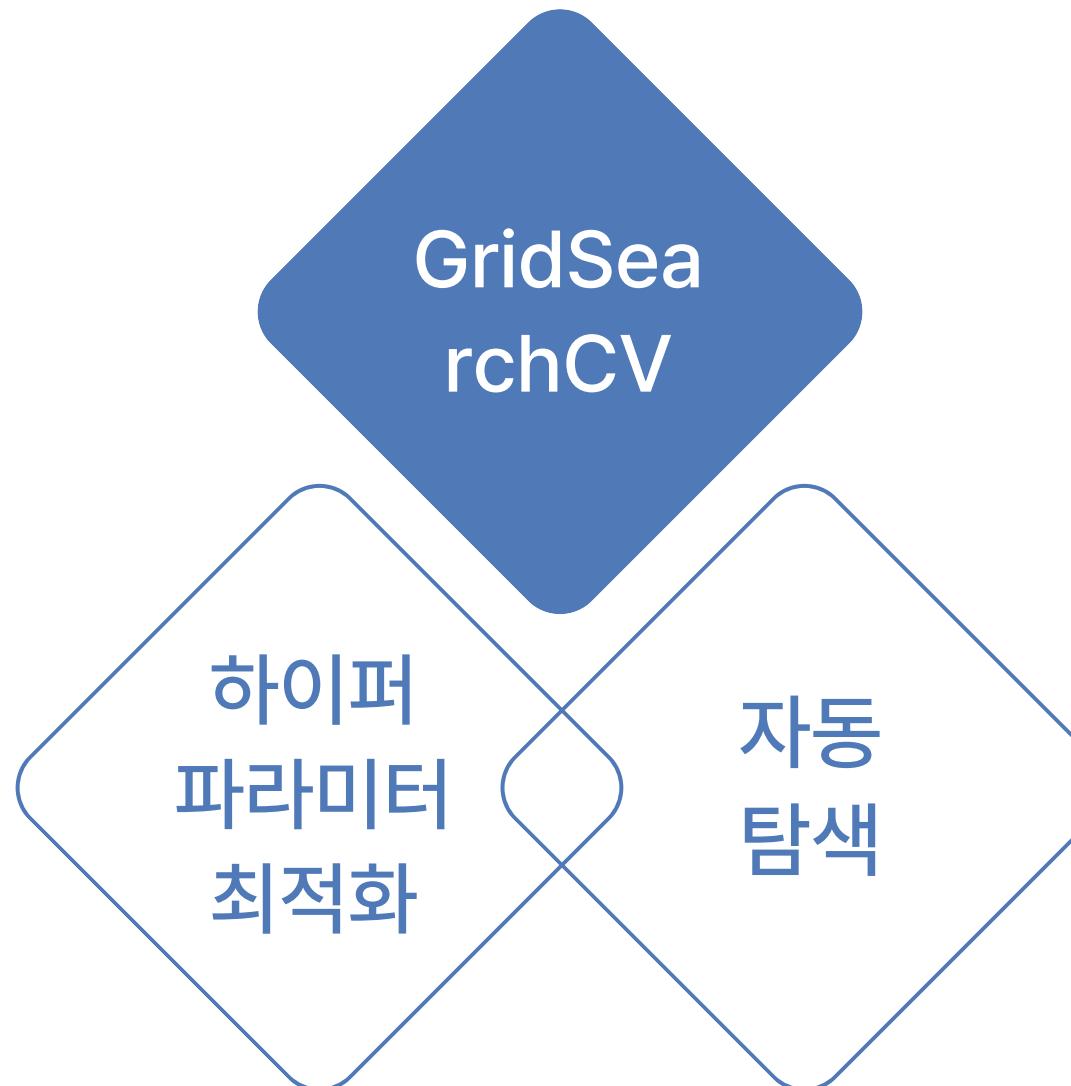


- 교차 검증을 위해 TimeSeriesSplit 사용
- 데이터를 시간 순서대로 5개의 분할로 나눈다.
- 각 분할에서 이전 데이터로 학습하고 다음 시간 구간의 데이터로 검증
- 이 과정을 통해 시계열 데이터의 특성을 고려하면서 모델의 일반화 성능 개선

03 모델링

-4. 개선 방안

교차 검증

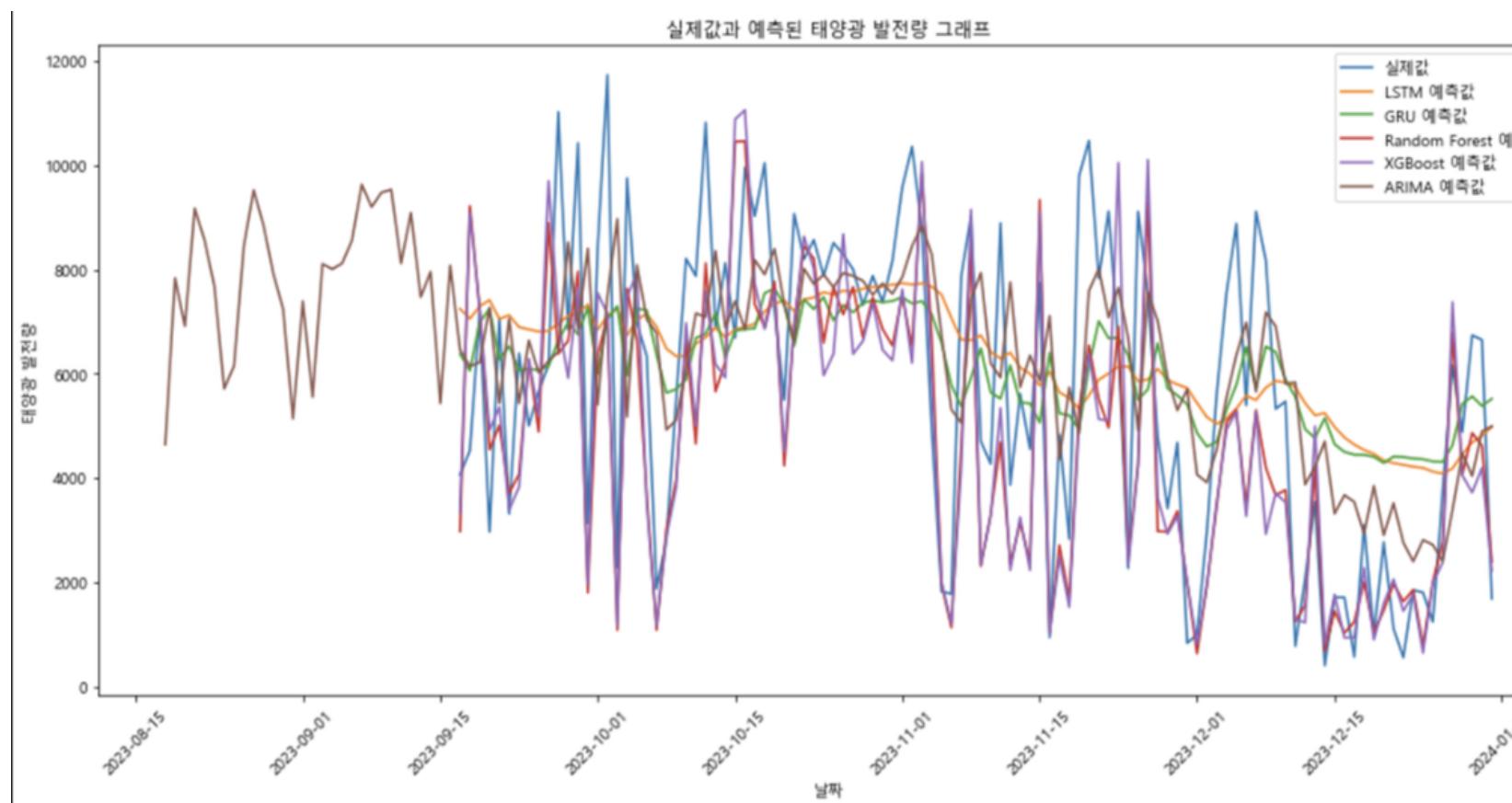


- 여러 하이퍼파라미터 조합을 체계적으로 시도
- Random Forest와 XGBoost 모델에 대해 n_estimators, max_depth 등의 파라미터 범위를 정의
- GridSearchCV를 사용하여 각 조합에 대해 교차 검증을 수행하고 성능을 평가
- 이를 통해 최적의 하이퍼파라미터 조합을 자동으로 찾을 수 있었다.

03 모델링

-5. 개선 후 성능 평가

태양광 발전량 성능 개선 후 평가



	MSE	MAE	NMAE	R2	Accuracy
LSTM	7.236497e+06	2220.925302	92.523134	0.190026	19.002631
GRU	6.808531e+06	2160.566194	90.008590	0.237928	23.792813
Random Forest	4.123370e+06	1607.174712	66.954454	0.538475	53.847545
XGBoost	4.622550e+06	1710.878550	71.274727	0.482603	48.260271

모델 개선 전후 비교 결과, 전체적으로 성능이 향상되었으나
여전히 Random Forest 모델이 가장 우수한 성능을 보인다.

03 모델링

-5. 개선 후 성능 평가

결론

LSTM/GRU

태양광 발전량 예측에서
개선 후 성능이 향상되었으
나 풍력 발전량 예측에서는
상대적으로 낮은 성능을
보임

Random Forest

태양광과 풍력 발전량 예측
시 일관되게 가장 높은
성능을 보임

ARIMA

두 예측에서
상대적으로 낮은
성능을 보임

XGBoost

두 예측에서
상대적으로 높은
성능을 보임

태양광 및 풍력 발전량 예측에 있어
Random Forest 모델이 가장 신뢰할 만한
결과를 제공할 것이라 예상

- 시계열 데이터의 특성을 고려할 때, LSTM이나 GRU와 같은 딥러닝 모델의 지속적인 개선 및 최적화가 필요.
- 실제 적용 시 모델의 성능 뿐 아니라 해석 가능성, 계산 복잡성 등을 종합적으로 고려해야 함.

재생 에너지 발전량 예측의 정확도를 높이는 데 기여할 수 있으며,
이는 전력 계통 운영의 안정성과 효율성 향상에 도움이 될 것으로 기대된다.

04 기대효과 및 결론

-1. 기대효과 및 결론

신재생 에너지의 경제적 효과와 탄소 배출 절감 평가

[태양광 탄소 배출량]	
전체 실제 발전량	4007946.38
전체 예측된 발전량	3723911.20
발전량 차이	-284035.18
탄소 배출량 감소 비율	-1.19 %

[태양광 정산 단가 가격 이득]	
전체 실제 태양광 발전량 총 가격	686372438.56
전체 예측된 태양광 발전량 총 가격	649136044.85
총 태양광 가격 차이	-37236393.71

[풍력 탄소 배출량]	
전체 실제 발전량	175505276.27
전체 예측된 발전량	1759352288.02
발전량 차이	459011.75
탄소 배출량 감소 비율	0.04 %

[충력 정산 단가 가격 이득]	
전체 실제 풍력 발전량 총 가격	30187151244.30
전체 예측된 풍력 발전량 총 가격	30340381383.32
총 풍력 가격 차이	-37236393.71

제주시 모델 적용 효과

탄소 배출 감소

효율적인 에너지 관리

전력 비용 절감

태양광과 풍력 발전량 예측을 통해 에너지 관리 및 비용 절감 실현

예측된 발전량을 기반으로 탄소 배출 감소 및 연료원별
절약된 비용 정량적 평가

감사합니다.

Team 브다히어로