

# 기상정보 다각화 및 통합 모델을 통한 태양광 에너지 발전량 예측 기술 연구

하재열\*, 이보람\*\*

\*한전KDN

## Research on Solar Energy Generation Forecasting Technology through Diversification and Integration of Meteorological Information Models

Jaeyeol Ha\*, Boram Lee\*\*

\*Kepco KDN

### ABSTRACT

본 연구는 다양한 기상정보 수집 모델을 활용하여 태양광 발전량을 예측하는 기술 개발을 목표로 한다. 국내 태양광 발전소는 대규모 발전소의 건설과 정부의 발전량 예측 수익 개선 등 다양한 원인으로 발전량 예측이 점점 중요해지고 있다. 이에 국내외 기상예보 및 실측 데이터의 다각화를 통해 발전량 예측 오차를 줄이고, 다양한 기계학습 인공지능 모델을 적용하여 예측 정확도를 향상하는 방법을 제안한다. 특히, 랜덤 포레스트, LightGBM, XGBoost 등 다양한 알고리즘을 비교 분석하고, 앙상블 기법을 통해 개별 모델의 한계를 극복하였다. 연구 결과, 제안된 앙상블 모델은 기존 단일 모델 대비 NMAE (Normalized Mean Absolute Error) 기준 13% 이상의 예측 정확도 향상을 보였다. 이를 통해 계통 운영의 안정성 증대, 에너지 관리 효율화, 그리고 재생에너지 정책 결정에 유용한 자료를 제공할 수 있을 것으로 기대된다. 궁극적으로 본 연구는 재생에너지 발전량 예측제도의 고도화와 디지털·그린 경제 활성화에 기여하고자 한다.

### 1. 서론

현대 사회에서 에너지 문제는 전 세계적으로 중요한 관심사로 떠오르고 있으며, 특히 재생 가능 에너지는 환경 보호와 지속 가능한 발전을 위한 핵심 요소로 자리 잡고 있다. 그중 태양광 에너지는 높은 잠재력과 청정성 덕분에 주요 재생 에너지 원으로 주목받고 있다. 그러나 태양광 에너지는 기상 조건에 큰 영향을 받기 때문에, 기상 정보를 정확하게 활용하는 것이 에너지 관리 및 예측의 효율성을 높이는 데 매우 중요하다.

현재 대부분의 태양광 발전량 예측 모델은 일사량이나 온도와 같은 단일 기상 정보에 의존하고 있다. 하지만 이러한 접근법은 기상 조건의 복잡성과 변동성을 충분히 반영하지 못하며, 기상 예측 오차에 민감하다는 한계가 있다.

본 연구에서는 다양한 기상 정보 수집 경로를 활용하여 태양광 발전량을 보다 정확하게 예측할 수 있는 방법을 제안하고자 한다. 특히 국내 태양광 발전량 예측 제도와 정책에 맞춰 예측 정확도를 개선하고, 이를 통해 에너지 운영 효율성을 높이며, 궁극적으로 에너지 관리 및 정책 결정에 유용한 자료를 제공하는 것을 목표로 한다.

### 2. 본문

#### 2.1 데이터 수집

본 연구를 위해 국내외 다양한 기상 사이트로부터 기상예측 정보를 가져와 태양광 발전량 실측 데이터와 비교하여 발전량 예측 정보를 도출하고자 하였다.

##### 2.1.1 국내 및 해외 기상예보 수집

데이터는 국내 기상청의 기상예측 자료를 기반하고 있으며, 해외 기상사이트 및 태양광 발전량 실측 센서 등 여러 사이트를 기반으로 발전량 예측과 결과를 비교해보고자 한다. 전체 데이터는 2023년1월1일부터 2024년 8월31일까지 시간 데이터로 구성되며 사용한 항목은 <표 1>과 같다. 회귀분석 검증을 위해 수집된 데이터의 2023년1월1일부터 2024년 6월31일까지 학습에 사용하였고, 나머지 2024년 8월 31일까지를 테스트에 사용하였다.

표 1 국내외 수집 데이터 항목

Table 1 Domestic and international collection data items

수집처	수집 데이터(변수)
국내기상예보	시간, 기온, 날씨정보, 일출시간, 일몰시간, 등
해외기상예보 - 1	시간, 기온, 날씨정보, 바람정보 등
해외기상예보 - 2	시간, 기온, 날씨정보, 구름정보 등
실측장치 및 기타 관측자료	시간, 실 발전량, 기온, 바람정보 등

날씨정보는 맑음, 흐림 등 정성적 수치로 나타나며 “맑음, 화창, 아이콘 형태의 맑음 표시” 등 예보 수집처별 각기 다른 형식으로 표기하고 있다. 이러한 데이터 특성을 가지는 변수들은 전처리 작업을 통해 통일하였다. 이후 범주형 데이터를 원핫인코딩(One-hot Encoding)을 통해 머신러닝에 적용할 수 있도록 변형하였다.

##### 2.1.2 데이터 통계분석

데이터 통계 회귀분석 결과는 <표 2>와 같다. R-squre의 경우 0.363으로 설명력은 높지 않으나 F-통계량의 p-value가  $2.30e-54$ 로 통계적으로 유의미한 변수를 가지고 있다. ADF test의 경우에도 0.0062로 시계열 데이터로 정상성을 띄고 있다. 등분산성의 경우  $2.08e-58$ 로 이분산성이 존재할 수 있다.

표 2 통계분석 지표

Table 2 Statistical analysis indicators

통계지표	상세	값	비고
$R^2$	R-squared:	0.3627	데이터 설명력
F검정	F-test	0.0000	유의수준 있음
정상성	ADF	0.0062	정상성 있음
정규분포	Shapiro-Wilk	0.0000	정상성 없음
자기상관	Durbin-Watson	0.5104	자기상관 있음
등분산성	Breusch-Pagan	0.0000	이분산성 있음

### 2.1.3 회귀모형 도출

회귀모형 비교를 위해 먼저 수집처 별로 발전량 예측을 진행했다. 국내 및 해외 기상예보를 각각 머신러닝 알고리즘 기법인 랜덤포레스트, XGBoost 및 LightGBM을 사용하여 평가를 진행하였다. 또한 국내와 해외 예보 데이터를 통합하여 예측을 진행하였다. 평가 산식은 태양광 발전량 예측에서 주로 사용하는 NMAE(Normalized Mean Absolute Error)로 진행하였다. NMAE가 낮을수록 성능이 좋은 알고리즘으로 볼 수 있다.

2024년 7~8월에 대한 모델 예측 결과는 <표3>과 같다. 국내 및 해외기상예보 단일모델의 경우 6.6~8.9% 정도의 오차율을 보이고 있으며 알고리즘 별로는 XGBoost가 가장 우수한 성능을 보였으며, 그 뒤를 LGBM과 랜덤 포레스트가 따르는 것으로 나타났다.

국내와 해외기상예보를 합친 통합 모델의 경우 5.8~6.5%의 오차율을 보이고 있으며 이 역시 XGBoost가 가장 우수한 성능을 보이고 있다. 통합 모델과 단일 모델간 비교해보면 각 모델의 최소 오차율을 기준으로 통합모델이 약 13% 이상의 예측 성능 개선을 보였다.

표 3 회귀모형 예측 결과

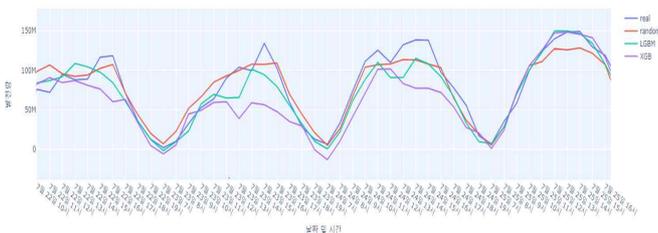
Table 3 Regression model prediction results

수집처	알고리즘	랜덤 포레스트	LGBM	XGBoost
	단일 모델	국내기상예보	8.9041%	8.7616%
	해외기상예보1	7.6016%	7.3548%	6.6345%
	해외기상예보2	7.8924%	7.5910%	7.3034%
통합 모델		6.5757%	5.9627%	5.8055%

<그림 1>에서 단일 기상 예보 모델을 활용한 예측 결과, 실제 발전량과 예측된 발전량 사이에 큰 편차가 발생하여 예측 정확도가 낮은 것으로 나타났다. 이는 단일 기상 정보에 기반한 예측 모델의 한계를 보여준다.

그림 1 국내기상예보(단일 모델) 예측 결과

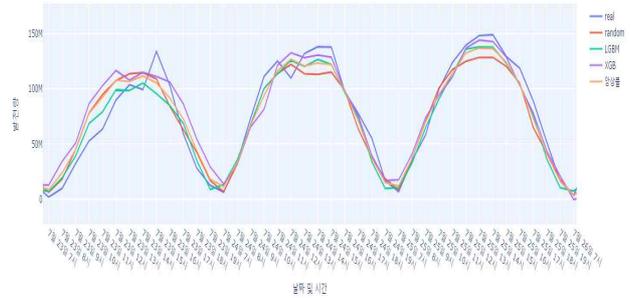
figure 1 Domestic weather forecast prediction results



반면, <그림 2>에서 통합 모델을 사용한 예측 결과에서는 실제 발전량과 예측된 발전량의 편차가 비교적 적어, 예측 정확도가 더 높은 것을 확인할 수 있다. 특히 통합 모델에서 사용된 알고리즘 성능 비교 결과, XGBoost가 가장 우수한 성능을 보였으며, 그 뒤를 LGBM과 랜덤 포레스트가 따르는 것으로 나타났다.

그림 2 통합 모델 예측 결과

figure 2 Integrated model prediction results



## 3. 결론

태양광 재생에너지 발전량 예측에서 기상 데이터의 수집 및 가공은 매우 중요한 과정이다. 기상 예보의 정확도가 낮을 경우, R-squared 지표가 함께 낮아져 모델의 설명력과 예측력이 저하될 수 있다. 이는 발전량 예측에 있어 중요한 요소로 작용하며, 기상 데이터의 품질과 다양성에 따라 예측 모델의 성능이 크게 달라질 수 있음을 의미한다. 본 연구에서는 기상 데이터의 다각화를 통해 다양한 기상 요인이 태양광 발전량에 미치는 영향을 분석하고, 예측 결과에서 유의미한 관계를 확인하였다. 이러한 분석은 단순한 기상 요인뿐만 아니라 복잡한 기상 조건을 고려한 예측 모델이 보다 높은 정확도를 보일 수 있음을 시사한다.

향후 연구에서는 기상 데이터의 품질을 높이고, 다양한 데이터 소스를 활용하여 보다 많은 양의 기상 데이터를 수집할 예정이다. 이러한 모델은 에너지 운영 효율성을 제고하고 태양광 발전 산업에 유용하게 활용될 수 있을 것이다.

## 참고 문헌

- [1] 심규연, 김동준, 정준홍, “기상정보를 이용한 태양광 발전량 예측서비스 개발,” 대한전기학회 학술대회 논문집, pp.2221-2222, July 2021.
- [2] Sung-Hyen Choi and Jin Hur, “Optimized-XGBoost Learner Based Bagging Model for Photovoltaic Power Forecasting,” The Transactions of the Korean Institute of Electrical Engineers, vol. 69, no. 7, pp. 978-984, July 2020.
- [3] Sangkyung Sung and Youngsang Cho, “Prediction of Photovoltaic Power Generation Based on Machine Learning Considering the Influence of Particulate Matter,” Environmental and Resource Economics Review, vol. 28, no. 4, pp. 467-495, December 2019.