

# 태양 위치 정보를 고려한 AutoML 기반의 태양광 발전량 예측

오진영<sup>1</sup>, 소다영<sup>2</sup>, 이병천<sup>1</sup>, 문지훈<sup>3</sup>  
<sup>1</sup>순천향대학교 AI·빅데이터학과 학부생  
<sup>2</sup>순천향대학교 ICT 융합학과 석사과정  
<sup>3</sup>순천향대학교 AI·빅데이터학과 교수  
 {wlsdud8261, sodayeong, qudcjs0208, jmoon22}@sch.ac.kr

## Automated Machine Learning-Based Solar PV Forecasting Considering Solar Position Information

Jinyeong Oh<sup>1</sup>, Dayeong So<sup>2</sup>, Byeongcheon Lee<sup>1</sup>, and Jihoon Moon<sup>1,2</sup>  
<sup>1</sup>Department of AI and Big Data, Soonchunhyang University  
<sup>2</sup>Department of ICT Convergence, Soonchunhyang University

### 요 약

지속 가능한 에너지인 태양광 발전은 전 세계에서 널리 활용하는 재생 에너지 원천 중 하나로 최근 효율적인 태양광 발전 시스템 운영을 위해 태양광 발전량을 정확하게 예측하기 위한 연구가 활발히 진행되고 있다. 태양광 발전량 예측 모델을 구성하기 위해서는 기상 및 대기 환경을 넘어 태양의 위치에 따른 일사량의 정보가 필수적이나 태양의 실시간 위치 정보를 입력 변수로 활용한 연구가 부족한 실정이다. 그리하여 본 논문에서는 시간과 태양광 발전소 위치를 기반으로 태양의 고도와 방위각을 실시간으로 계산하여 입력 변수로 사용하는 방식을 제안한다. 이를 위해 AutoML 기반의 다양한 기계학습 모델을 구성하여 태양광 발전율을 예측하고 그 성능을 비교 분석하였다. 실험 결과, 태양 위치 정보를 포함한 경우에 환경 변수만을 고려하였을 때보다 예측 성능이 크게 향상되었음을 확인할 수 있었으며, Extra Trees 모델의 경우 태양 위치 정보를 추가하였을 때 MAE (Mean Absolute Error)가 33.90 에서 22.38 까지 낮아지는 결과를 확인하였다.

### 1. 서 론

태양광(Photovoltaics) 발전은 지속 가능한 에너지 원천력으로서 세계 각국이 적극적으로 활용하고 있는 재생 에너지 원천 중 하나로 태양광 발전 시스템의 효율성과 안정성을 높이기 위한 연구가 활발히 진행되고 있다[1]. 특히, 정확한 태양광 발전량 예측은 전력 시스템의 안정적 운영과 에너지 관리에 필수적이기 때문에, 최근 시계열 예측에 강건한 인공지능 방법론을 이용한 연구가 활발히 진행되고 있다[2,3].

일반적으로 태양광 발전량은 기상 및 대기 환경의 영향을 크게 받기 때문에 이러한 변수들을 활용하여 발전량을 예측한다[4]. 예를 들어 습도, 일사량 등과 같은 기상 요인은 태양광 발전량에 있어 주요 변수로 활용된다. 여기서 태양 위치 정보는 태양광 발전량에 많은 영향을 미친다는 연구결과가 보고되고 있지만[5], 국내 태양광 발전량 예측 모델 구성에 사용된 연구는 극히 제한적이다.

따라서 본 논문은 태양 위치 정보를 태양광 발전량 예측 입력 변수로 활용하는 방법을 제안한다. 태양 위치 정보는 태양의 고도와 방위각에 대한 데이터를 의미하며, 본 연구에서는 선행 연구[6]를 고려하여 입력 변수로 구성하였다. 태양의 위치 정보가 최종 예측 성능에 미치는 영향을 확인하고자 위치 정보를 활용한 모델과 활용하지 않은 모델 간의 예측 성능을 비교하였으며, 이를 위해 AutoML (Automated Machine Learning) 기반의 GBM (Gradient Boosting Machine), ET (Extra Trees), RF (Random Forest), LightGBM (Light GBM), LR (Linear Regression), DT (Decision Tree) 등 다양한 인공지능 방법론을 이용하였다.

본 논문의 나머지 구조는 다음과 같다. 2 장에서는 데이터 셋 구성에 대하여 설명하고, 3 장은 제안하는 실험 모델에 대해 설명한다. 4 장에서는 다양한 예측 모델들에 대한 성능을 비교분석한다. 끝으로 5 장은 본 연구의 결론 및 향후 연구 방향을 제시한다.

2. 데이터 구성

본 논문은 태양광 발전량 예측 모델을 구성하고자 부산 신항 지역에서의 2013 년 1 월 1 일부터 2020 년 7 월 31 일까지 1 시간 단위로 기록된 기상 데이터와 태양광 발전량 데이터를 사용하였다. 표 1 과 같이 기상 데이터는 평균 기온, 일 최저 기온/최고 기온, 온도, 강수량, 일사량, 운량 등 14 개와 태양의 위치 정보 2 개로 총 16 개를 입력 변수로 사용하였다.

<표 1> 입력 변수 목록

Input Variables
<i>Average Temp, Lowest Temp, Highest Temp, Rain Fall, Steam Press, Dew Point, Sunshine, Insolation, Cloudiness, Ground Temp, Temp, Wind, Press, Humi, Solar_azimuth, Solar_altitude</i>

일반적으로 태양광 발전량 데이터는 태양광 발전 시설마다 설비용량이 다르므로 기록된 발전량 값은 정량적인 비교에 한계가 있다. 따라서 식 (1)에 따라 태양광 발전량 데이터를 각 발전소의 설비용량으로 나누어 발전율 데이터를 출력 변수로 활용하였다.

$$\text{발전율} = \text{발전량} / \text{설비용량} \quad (1)$$

최종적으로 데이터 표준화(Standardization)를 적용하여 16 개의 입력 변수와 1 개의 출력 변수를 구성하였다.

3. 실험 모델 구성

AutoML [7]은 기계학습 절차를 자동화하는 도구로, 데이터 관련 비전문가도 쉽게 기계학습 모델을 구축할 수 있다. 또한, AutoML 은 다양한 모델과 최적의 초매개변수를 적용하여 모델의 정확도를 향상할 수 있으며, 모델의 생성 과정을 투명화하여 모델의 설명 가능성을 높이는 유용한 도구이다.

본 논문에서 AutoML 도구를 활용하여 2019 년까지 데이터를 학습 집합, 2020 년 이후를 평가 집합으로 설정하여 실험을 수행하였다. 또한, 성능 평가 지표로 MAE (Mean Absolute Error)와 RMSE (Root Mean Square Error)를 사용하였다.

4. 실험 결과

표 2 는 각 모델에 관해 태양 위치 정보 데이터를 포함(Input Variables w/ Solar Position)하여 실험한 것과 데이터를 포함하지 않고(Input Variables w/o Solar Position) 실험한 결과를 나타내며 MAE 와 RMSE 를 평가 지표로 사용하였다.

대다수의 모델에서 태양 위치 정보를 포함하였을 때 태양광 발전량 예측 성능이 더 우수하다는 것을 확인하였다. 구체적으로 나무 기반 앙상블 기법들이 대체로 우수한 예측 성능을 도출하였으며, 태양 위치 정보가 포함되었을 때 뚜렷한 성능 향상을 보였다.

<표 2> 태양 위치 정보 적용 유무에 관한 실험 결과

Machine Learning (ML) Methods	Input Variables w/ Solar Position		Input Variables w/o Solar Position	
	MAE	RMSE	MAE	RMSE
Decision Tree	<b>24.537</b>	<b>63.177</b>	26.114	66.827
Random Forest	<b>22.878</b>	60.856	25.588	59.819
Extra Trees	<b>22.389</b>	<b>60.085</b>	33.906	62.768
Linear Regression	31.910	<b>65.771</b>	31.757	66.621
Gradient Boosting Machine (GBM)	<b>27.960</b>	<b>68.341</b>	33.020	73.793
LightGBM	<b>26.737</b>	<b>65.389</b>	29.067	68.790

5. 결론

본 논문은 AutoML 기반 다양한 기계학습 모델을 대상으로 태양 위치 정보를 새로운 입력 변수로 활용하여 태양광 발전율의 예측 성능을 태양 위치 정보를 반영하지 않은 모델과의 성능을 비교 및 분석하였다. 실험 결과 태양 위치 정보를 활용하면 예측 성능이 전반적으로 크게 향상할 수 있음을 확인하였다.

향후 본 논문에서 제안한 기법을 바탕으로 더 많은 지역의 태양광 발전량 예측을 수행할 예정이며, 단일 기계학습 모델을 활용하는 것을 넘어 앙상블, 최적화 기법 등을 통해 강건한 예측 모델을 구현할 것이다.

사 사 문 구

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구 결과로 수행되었음(2021-0-01399).

참 고 문 헌

- [1] S.-H. Cheon, “국내 신재생에너지정책 추진현황과 향후 전망”, *The Magazine of the Society of Air-Conditioning and Refrigerating Engineers of Korea*, Vol. 39, No. 1, pp. 6–12, 2010.
- [2] J.-J. Song et al., “Analysis of prediction model for solar power generation,” *Journal of Digital Convergence*, Vol. 12, No. 3, pp. 243–248, 2014.
- [3] W. Chung et al., “SHAP based Solar Power Generation Forecasting Scheme Reflecting Time Patter,” in *Proc. of the KCC 2021*, 2021, pp. 262–264.
- [4] J. H. Jeong and Y.-T. Chae, “Improvement for forecasting of photovoltaic power output using real time weather data based on machine learning,” *Journal of The Korean Society of Living Environmental System*, Vol. 25, No. 1, pp. 119–125, 2018.
- [5] J. Lee et al., “Analysis on the Electricity of Photovoltaic Systems by Applying the Contribution Factor of Solar Light Elements to Power Generation,” in *Proc. of the KIEE Conference*, 2010, pp. 56–58.
- [6] M.-H. Seo et al., “A Development of Solar Position Algorithm for Improving the Efficiency of Photovoltaic Power Generation,” In *Proc. of KIIT Conference*, 2009, pp. 46–51.
- [7] X. He et al., “AutoML: A survey of the state-of-the-art,” *Knowledge-Based Systems*, Vol. 212, p. 106622, 2021.