

# 머신러닝 기반 태양광 에너지 발전량 예측 및 모델별 비교 연구

이보람, 하재열, 이태영  
한전 KDN

## A Study on the Prediction of Solar Energy Generation Based on Machine Learning and Comparison of Forecasting Models

Lee BoRam, JaeYeol Ha, TaeYoung Lee  
Kepco KDN

### ABSTRACT

재생에너지 입찰제도의 예측 보상 및 패널티 제도를 대상으로 AI 기술을 이용한 신재생에너지 발전량 예측 기술의 중요성이 부각되고 있다. 태양광 발전량을 정확하게 예측하는 AI 모델을 찾기 위해 본 연구에서는 머신러닝 기반의 알고리즘을 활용하여 예측 모델을 학습하고 모델별 성능을 비교하였다. 기상 예보사이트에서 얻은 예보데이터를 학습데이터로 활용하였으며, RandomForest, LightGBM, XGBoost, 세가지 머신러닝 알고리즘으로 각각의 모델을 학습하였다. KPX의 재생에너지 예측제도에서 활용하는 성능평가 기준을 사용하여 모델별 오차를 구하고, 성능을 비교하였다. 시뮬레이션 결과 XGBoost가 가장 좋은 성능을 가지고 있다는 것을 확인할 수 있었다.

### 1. 서론

최근 재생에너지 보급이 증가하면서 전력계통의 불안정성이 높아지고 있다. 전력 수요에 대응해 전력 계통의 발전량을 맞추고 불안정성을 보완하기 위해 재생에너지 입찰제도의 예측 보상 및 패널티 제도가 도입되었다. 하루 전 오전 오후에 에너지 시장에 입찰하는 방식이며 이와 같은 제도의 도입으로 인해 다음날의 발전량을 정확하게 예측하는 기술의 필요성이 대두되고 있다. 본 논문에서는 국내 태양광 발전사 중 한 곳에서 제공 받은 발전량 데이터와 기상예보데이터를 전처리하고 머신러닝 알고리즘을 활용하여 태양광 발전량을 예측하는 모델을 생성한 뒤 그 결과를 비교할 예정이다. 본문에서는 데이터 전처리방안과 태양광 발전량 예측 모델을 생성하는 방안에 대해 기술하고 그 결과를 평가하였으며, 결론에서는 생성된 모델을 비교하는 내용으로 마무리 짓는다.

### 2. 본론

#### 2.1 태양광 발전량 예측 모델

머신러닝을 활용한 태양광 발전량 예측 기술에 대한 다양한 연구가 진행되어 왔지만, 당일 실측 데이터를 활용하여 그날의 발전량을 예측하는 연구가 대부분이다. 현재 재생에너지 입찰제도에서는 전날 오전 10시와 오후 5시에 시간별 발전량 예측값을 제출해야 하며, 오차율이 8% 이내여야 입찰제도를 이용할 수 있다. 재생에너지 입찰제도에 참여하기 위해서 당일이 아닌 전날까지의 데이터를 이용해 발전량을 예측하는 기술이 필요하다.

본 논문에서는 전날 예보데이터를 활용해 태양광 발전량을 예측하는 모델을 다양한 알고리즘으로 학습하고, 이를 비교 분석하였다. 2.2절에서는 전반적인 연구 방법에 대해 소개하였다. 2.2.1절에서는 태양광 발전량 예측에 활용된 데이터 및 전처리 방법에 대해 설명하였고, 2.2.2절에서는 예측 모델을 학습한 알고리즘에 대해 설명하였으며, 2.2.3절에서는 각 모델에 따른 결과를 평가하며 마무리 짓는다.

#### 2.2 연구 방법

##### 2.2.1 활용 데이터셋

태양광 발전량에 영향을 미치는 요소는 대표적으로 일사량, 일조량 등이 존재한다. 하지만 기상 예보 사이트에서 제공하는 데이터는 한계가 있어, 기상 및 기온, 일출, 일몰 시간에 대한 예보를 활용하여 학습을 대신할 예정이다. 기상청과 해외 기상 예보 사이트에서 제공하는 API를 활용하여 전날 예측한 기상 예보 데이터를 확보하였고, 국내 태양광 발전회사의 발전량 데이터를 확보하여 학습데이터를 구축하였다. 총 1년간의 데이터를 확보하였으며, 데이터 중 랜덤하게 80%를 추출하여 학습데이터에 사용하였고, 나머지 20%의 데이터는 테스트데이터로 사용하였다. 먼저 API를 통해 예보데이터를 데이터베이스에 저장하고 기상 데이터별 빈도수를 확인하여 빈도수가 낮고 '빗방울', '비'와 같이 과하게 세분화 된 기상 데이터를 하나의 기상으로 합쳐주는 전처리 과정을 진행하였다. 전처리가 끝난 범주형 데이터는 원-핫 인코딩을 통해 수치형 데이터로 변환하여 머신러닝 학습에 적합하게 만들었다.

##### 2.2.2 머신러닝 알고리즘

본 논문에서는 RandomForest, LightGBM, XGBoost 세가지 머신러닝 알고리즘을 이용하여 각각 학습을 진행했고 결과를 비교 분석하였다.

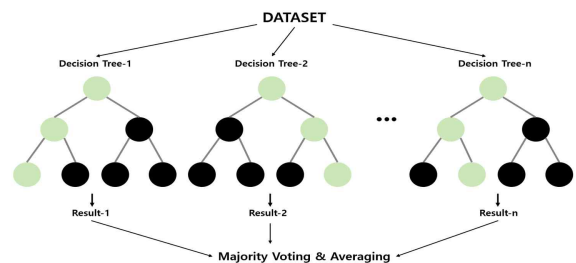


그림 1 랜덤포레스트 기법

RandomForest는 랜덤하게 여러 개의 의사결정나무를 만들고 각각의 예측을 취합하여 최종 예측을 하는 앙상블 기법을 사용하며, 각 나무가 독립적으로 학습하여 과적합 방지에 효과적인 알고리즘이다.

LightGBM은 의사결정 나무를 만들 때, 오차를 줄이는 방향으로 학습하는 GBM기법을 활용하며, 병렬 학습이 가능하여 속도가 빠르다는 장점이 있는 알고리즘이다. 특히 큰 데이터를 다룰 때 효율적이나, 데이터가 적을 경우 과적합의 위험성이 높은 알고리즘이다.

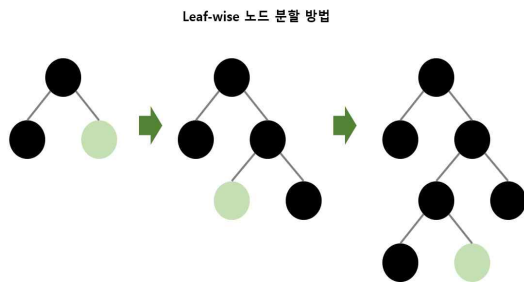


그림 2 LightGBM 의사결정나무 생성 방법

XGBoost은 여러 개의 의사결정나무를 순차적으로 만들고, 이전 나무의 오류를 보완하는 방식으로 학습하는 알고리즘이다. 병렬 학습이 가능해 학습 속도가 빠르고 과적합을 방지하는 기법을 추가해 GBM의 단점을 보완한 머신러닝 기법이다.

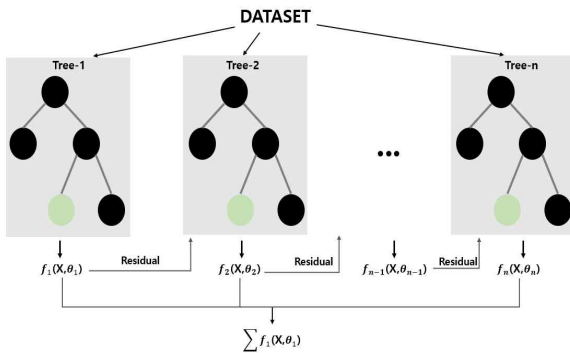


그림 3 XGBoost 알고리즘

### 2.2.3 학습 결과

학습에 대한 평가지표는 KPX의 재생에너지 발전량 예측제도에 활용되고 있는 예측오차율 식(1)을 사용하였다.

$$\text{예측오차율}(\%) = |\text{발전량} - \text{예측량}| / \text{발전량} \times 100(\%) \quad (1)$$

테스트 결과 Random Forest의 예측오차율은 6.79%, LightGBM은 6.31%, XGBoost는 6.08%의 오차율을 확인할 수 있었으며 XGBoost의 성능이 가장 좋은 것을 확인할 수 있었다. 표1은 모델별 평가 결과를 정리해 놓은 것이며 세 모델 전부 재생에너지 입찰제도를 이용할 수 있는 오차율 8% 이내의 성능을 가지고 있는 것을 보여주고 있다.

표 1 모델별 평가 결과

model	RandomForest	LightGBM	XGBoost
예측오차율(%)	6.7992	6.3121	6.0309

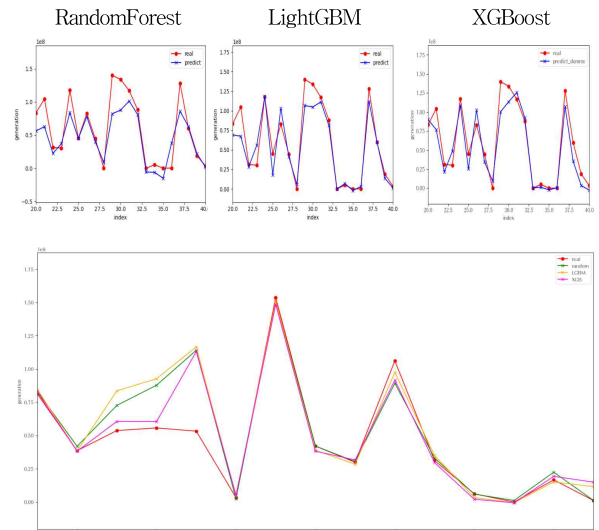


그림 4 세가지 모델 비교 시각화

### 3. 결론

본 논문에서는 예보데이터를 이용해 태양광 발전량 예측하는 모델을 학습해보고 KPX에서 이용하는 오차율 식을 이용해 그 결과를 비교하였다. RandomForest, LightGBM, XGBoost 세 가지 머신러닝 알고리즘을 사용하여 모델을 생성하였으며 각 모델별 오차율은 약 6%로 전부 재생에너지 입찰제도를 이용할 수 있는 8% 이내의 낮은 오차율을 보여주었다. 그 중 XGBoost모델은 오차율 6.08%로 가장 우수한 결과를 보여주었으며, 일부분을 시각화한 그림을 확인해보았을 때도 XGBoost의 성능이 더 좋은 것을 확인할 수 있었다. 이번 연구를 통해 태양광 예보데이터를 이용하여 만든 태양광 발전량 예측 모델의 성능을 확인하였으며, 그 중 XGBoost 알고리즘을 이용하여 만든 모델의 성능이 가장 우수한 것을 확인할 수 있었다.

향후 태양광 발전사에 예측 모델을 적용하는 실증을 진행하고 그 결과를 바탕으로 오차분석을 통해 재학습하는 연구가 이루어진다면, 태양광 발전량 예측에 최적화된 모델을 생성할 수 있을 것이라 기대된다.

### 참고 문헌

- [1] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and regression trees. Routledge, 2017.
- [2] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," Advances in neural information processing systems, vol. 30, 2017.
- [3] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785 - 794, 2016.