

# 랜덤포레스트 모델을 이용한 에너지 다소비 식품공장의 에너지 수요량 예측

김동주, 이형아, 구재희

고등기술연구원 에너지환경IT융합그룹

kimdongju@iae.re.kr, lhaeve@iae.re.kr, jaehoi@iae.re.kr

## Prediction of energy demand in energy heavy consuming food factory using random forest model

Dongju Kim, Hyungah Lee, Jae-Hoi Gu

Institute for Advanced Engineering, Energy environment IT convergence group

### 요 약

본 논문은 식품공장의 에너지 사용량 절감을 위한 공장에너지관리시스템(FEMS)에 적용하고자 머신러닝 기반 랜덤포레스트 알고리즘을 사용하여 에너지 소비량 예측 모델을 도출하였다. 수집된 제품 총생산량과 도시가스 사용량 데이터를 기반으로 독립변수 확장 전처리 기법을 적용하여 84개의 독립변수로 파생하였으며, 데이터마이닝 도구인 이썬마이너에서 제공하는 랜덤포레스트 알고리즘의 초매개변수 중 하나인 총 나무수 값을 조정하여 모델을 구성하였다. 본 연구를 통해 식품공장에서 도시가스 에너지의 수요량을 예측할 수 있는 최적 성능의 모델을 개발하였다.

### I. 서 론

2021년 기준 국내 식품산업의 생산실적 매출액은 약 72조 4백억원 수준으로 전년 대비 약 9.6 % 증가하였으며, 이는 2015년 이래 지속적인 증가 추세를 보인다 2021년에 들어 비약적으로 증가한 것이다. 또한 국내 제조업 총생산(GDP) 대비 11.45% 수준으로 국내 산업 분야에서 식품 산업의 규모가 점점 증가하고 있음을 알 수 있다.[1]

식품산업은 대표적인 에너지다소비 업종으로서 정부의 2050년 탄소중립 목표 달성을 위한 에너지 절감 및 온실가스 배출량 저감이 요구되고 있으며, 3차 에너지기본계획에 의해 2025년부터 10만TOE 이상의 에너지를 소비하는 사업장을 대상으로 공장에너지관리시스템(FEMS) 설치의무화 시행이 예정되어 있어 에너지 모니터링뿐만 아니라 정밀한 에너지 수요예측을 기반으로 한 에너지 소비 최적화에 대한 요구가 증가되고 있다.[2][3]

공장에너지관리시스템은 ICT 기술과 수집된 데이터를 기반으로 한 제어기술을 활용하여 소비 에너지를 최적화 할 수 있는 기술이다. 기존에는 고효율의 설비 또는 장비의 도입을 통한 에너지 절감 사례가 대부분이었으나, 공장에너지관리시스템은 기존 장비의 에너지 사용현황, 설비의 가동현황을 모니터링하며 수집된 데이터의 분석, 예측을 통해 공급량을 제어함으로써 에너지 소비의 합리화를 도모할 수 있다.[2]

공장에너지관리시스템에서 에너지를 적재적소에 공급하기 위해서는 실시간으로 수집되는 데이터를 수용한 정밀한 수준의 에너지 수요량 예측이 요구된다. 또한, 대부분의 식품공장들과 같이 영세한 규모의 공장에서 수집될 수 있는 데이터의 종류가 제한적이기 때문에 최소한의 변수를 바탕으로 최적의 예측 성능을 도출하는 모델의 개발이 필요하다.

### II. 본론

본 연구에서는 열에너지를 주로 사용하는 식품가공공장의 공장에너지관리시스템(FEMS)을 통한 에너지 최적화를 달성하기 위해 에너지 수요량 예측 모델을 구축하였다. 식품가공공장에서 데이터를 수집하여 인공지능 기반의 머신러닝 기법을 적용한 도시가스 수요량 예측 모델을 구축

하고 모델 간 성능비교를 통해 최적 조건을 도출하는데 목적이 있다.

### II-1 데이터 수집 및 전처리

본 연구에서 사용한 원본 데이터는 육가공제품을 생산하는 식품생산공장에서 제공한 수기일지를 전산화 작업을 통해 획득하였다. 수집된 독립변수는 도시가스 사용량과 제품생산량이며, 수집기간 및 데이터 내용은 Table 1에 나타내었다. 제한된 독립변수의 수를 사용하여 예측 모델의 성능을 향상시키기 위해 당일부터 5일전까지의 도시가스 사용량과 제품 생산량을 개별의 변수로 지정하여 12가지의 변수로 나타내었다.

Variable index	Data contents (Collection period : 2017. 7. 3 ~ 2020. 10. 20)
1	Amount of LNG consumption today
2	Amount of LNG consumption one days ago
3	Amount of LNG consumption two days ago
4	Amount of LNG consumption three days ago
5	Amount of LNG consumption four days ago
6	Amount of LNG consumption five days ago
7	Total amount of product today
8	Total amount of product one days ago
9	Total amount of product two days ago
10	Total amount of product three days ago
11	Total amount of product four days ago
12	Total amount of product five days ago

Table 1. Description of raw collected data

수집된 12가지의 변수데이터는 최대-최소 정규화 전처리 기법으로 데이터 간 크기를 균일화 하였고, 식(1)과 같이 일주일 단위로 변수를 확장하여 84개의 독립변수와 1일 후의 도시가스 사용량으로 구성된 종속변수로 데이터 전처리를 완료하였다.[4]

또한, 모델의 학습과 모델의 성능 평가를 위해 수집된 전체 데이터셋을

학습 데이터셋과 테스트 데이터셋으로 구분하였으며, 학습 데이터셋은 2017.7.3. ~ 2019.12.30.까지의 901개 데이터를 사용하였고, 테스트 데이터셋은 2020.10.1. ~ 2020.10.20.까지의 294개 데이터를 사용하였다.

$$X_i = \text{concat}(x_i, x_{i-1}, x_{i-2}, x_{i-3}, x_{i-4}, x_{i-5}, x_{i-6}) \quad (1)$$

$$\text{where } x_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{i12}]^T$$

## II-2 이썬마이닝을 통한 랜덤포레스트 모델

전처리된 데이터를 바탕으로 에너지 수요량 예측 모델을 구성하기 위해 데이터마이닝 도구인 이썬마이닝을 사용하였다. 이썬마이닝에서 제공하는 알고리즘 중 랜덤포레스트를 사용하여 수요량 예측 모델을 구성하였다. 랜덤포레스트 모델은 의사결정나무 알고리즘의 결합 모델로서 일부 의사결정 나무가 잘못된 예측을 하더라도 다른 모델의 예측 결과를 종합하여 정확한 예측을 가능케 한다는 장점이 있다.[4]

본 연구에서 사용한 랜덤포레스트 모델의 초매개변수를 아래 Table 2에 나타내었다.

Hyperparameter	Value
Number of trees	10, 100, 300, 500, 1000, 3000, 10000
Max depth	0
Impurity function	Least Squared Deviation
OOB (Out of bag)	Used

Table 2. Hyperparameter of Random Forest which used

도출된 모델에 대한 예측 성능을 분석하기 위해 사용한 지표는 결정계수(R-square)와 평균 절대 비율 오차(MAPE)를 사용하였으며, 식 (2), (3)에 나타내었다.

$$R^2 = 1 - \frac{\sum_i^N (y^{(i)} - \hat{y}^{(i)})^2}{\sum_i^N (y^{(i)} - \bar{y})^2} \quad (2)$$

$$MAPE = \frac{100}{N} \sum_i^N \left| \frac{y^{(i)} - \hat{y}^{(i)}}{y^{(i)}} \right| \quad (3)$$

## II-3 변수 확장에 따른 모델 성능 비교

학습 데이터셋의 하루 전 생산량, 도시가스 소비량 변수를 기반으로 당일의 도시가스 소비량을 종속변수로 한 모델의 예측에서 데이터의 최대-최소정규화, 변수확장기법이 랜덤포레스트 모델의 예측성능에 미치는 결과를 아래 Table 3에 나타내었다.

(총 나무수 : 300)	R-square	MAPE
단일 변수 적용	0.24677	18.44015
최대-최소 정규화 적용	0.25134	18.30919
정규화 + 변수확장	0.79540	9.78681

Table 3. Model performance comparison of variable expansion dataset

수집된 두 개의 단일 변수 및 최대최소 정규화 기법만으로는 모델의 예측 설명력이 낮은 결과를 확인하였으며, 변수의 개수를 84개로 확장한 데이터를 활용한 모델은 결정계수가 0.795 수준으로 높아져 당일의 도시가스 소비를 예측하는데 충분한 설명력을 갖추는 것으로 확인되었다. MAPE 역시 단일 변수 및 정규화 기법적용 시 18.3 ~ 18.4 수준으로 확인

되었으나, 변수확장 적용 결과 9.796 수준으로 보다 높은 설명력을 갖추는 결과를 보였다.

## II-4 랜덤포레스트 에너지 수요 예측 모델 성능 평가

84개의 변수로 확장한 학습 데이터셋을 바탕으로 학습한 랜덤포레스트 모델에 동일한 데이터 전처리를 거친 테스트 데이터셋을 입력하여 도시가스 소비량 예측 결과를 분석하였다. 아래 Figure 1.에 랜덤포레스트 모델의 총 나무수를 300으로 설정한 모델의 예측 결과를 나타내었다. 도시가스가 사용되지 않는 주말과 공장이 정상적으로 가동되는 주중의 모델 분석 결과 설명력이 높은 예측 성능을 나타내는 것을 확인할 수 있으나, 일부 도시가스 사용량이 이상적으로 증가한 날짜에 대한 예측에는 한계가 있음을 확인하였다.

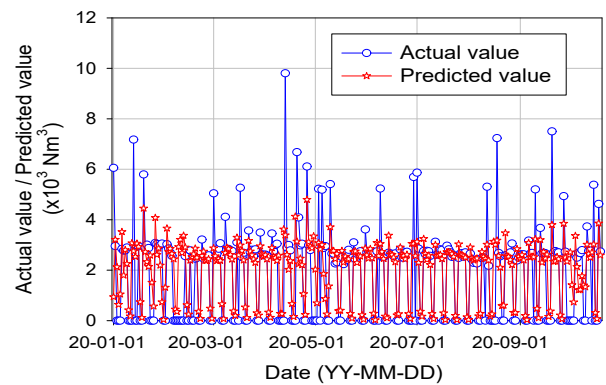


Figure 1. Comparison of actual value and predicted value using by Random Forest model (Tree 300)

랜덤포레스트 모델의 초매개변수 중 하나인 총 나무수의 값에 따른 영향을 확인하기 위해 총 나무수 값을 10 ~ 10,000까지 변경하여 테스트 데이터셋을 입력하였으며, 실제값 대비 예측값의 모델 성능을 결정계수와 MAPE로 평가하여 Figure 2.에 나타내었다. 학습 데이터셋에 비해 테스트 데이터셋을 입력한 모델의 성능은 대체로 낮아지는 경향을 확인하였으며, 랜덤포레스트 모델의 총 나무수의 값이 300까지 증가할수록 결정계수 및 MAPE 평가가 증가하는 모습을 나타내었다. 300 ~ 10,000개의 총 나무수를 입력한 모델의 경우 성능의 증가폭이 미미한 것으로 확인되었다.

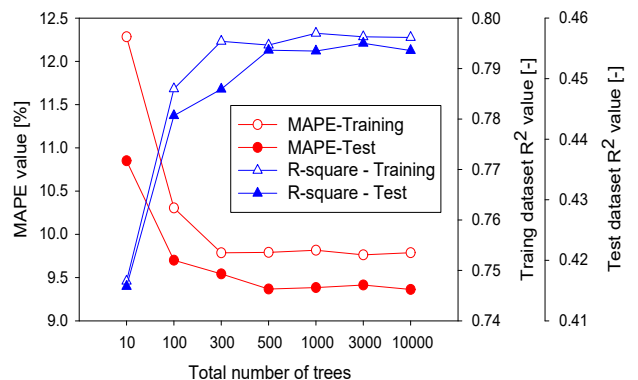


Figure 2. Comparison of model performance according to total number of trees

### III. 결론

본 논문에서는 대표적인 에너지 다소비 업종인 식품가공공장에서 수집한 데이터를 바탕으로 랜덤포레스트 모델을 적용하여 예상되는 에너지의 수요량을 예측하였다. 제한된 최소한의 변수를 활용하여 도시가스의 소비량을 예측하기 위해 변수 확장기법을 적용하였으며, 전처리를 거친 데이터셋으로 랜덤포레스트 모델을 훈련하여 예측 성능을 비교하였다. 수집된 상태의 데이터를 바탕으로 모델을 학습한 결과 투입한 변수가 예측결과를 설명하는데 부족함이 있었으며, 대표적인 전처리 기법인 최대-최소정규화를 통해서도 성능이 향상되지 못한 결과를 확인하였다. 그러나 변수 확장기법을 통해 증가된 변수들이 상호 보완작용을 통해 예측값의 설명력을 높인 것으로 확인되었으며, 이러한 성능 증가에 대해서 추가적인 검증이 필요할 것으로 사료된다.

또한, 본 연구에서 활용한 랜덤포레스트 모델의 초매개변수인 총 나무수의 변화에 따른 모델의 성능 비교 결과 300그루까지는 성능의 증가가 확인되었으나, 그 이상의 값에서는 성능의 증가가 미미한 것으로 확인되었다. 총 나무수의 증가는 모델의 분석 및 연산 시간과 직결적인 관계를 갖고 있어 입력값에 대한 최적점을 도출할 필요성이 있으며, 대상 식품공장의 도시가스 수요량을 예측하기 위해선 300그루의 총 나무수가 적합한 것으로 확인하였다.

랜덤포레스트 모델의 초매개변수 조정 및 학습 데이터셋의 과적합을 방지하기 위한 데이터 범위 조정을 바탕으로 추후 연구를 수행하여 보다 설명력 높은 모델을 구현할 수 있을 것으로 예상된다.

### ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구과제입니다. (No. 20202020800290)

### 참 고 문 헌

- [1] 2021 식품 등의 생산실적, 식품의약품안전처, 2022.07.
- [2] Dongju K., Chae-Eun Y., Jae-Hoi G., "Analysis of Food Factory Energy Consumption Pattern for Application of Factory Energy Management System(FEMS)", Journal of thermal environmental & engineers, 16(2), pp. 1-8, December 2021.
- [3] Youngmee S., Yoonmee D., Tae-Wook H., Il-woo L., "Common Framework for Factory Energy Management System(FEMS)", Proceedings of Symposium of the Korean Institute of Communications and Information Sciences, pp. 453-454, June 2022.
- [4] Jinsoo B., Seoung Bum K., "Predictions of COVID-19 in Korea Using Machine Learning Models", Journal of the Korean Institute of Industrial Engineers, 47(3), pp. 272-279, June 2021.