

트리 기반 알고리즘을 활용한 사무용 건물 에너지 소비량 예측모델 개발 및 변수 중요도 분석

Development of an Office Building Energy Consumption Prediction Model Using a Tree-Based algorithm and Analysis of Variable Importance

○김 혜 인* 윤 지 영** 김 준 규* 문 진 우***
Kim, Hye In Yun, Ji Young Kim, Jun Kyu Moon, Jin Woo

Abstract

The purpose of this study was to develop an annual energy consumption prediction model for office buildings using a tree-based ensemble machine learning algorithm and to evaluate its applicability by comparing the predicted results with actual building data. The prediction model was developed using the ComStock dataset provided by the U.S. Department of Energy, employing the Extra Trees Regressor algorithm. The model was trained on 55,385 building samples from the North Carolina 4A climate zone. The results demonstrated that the model achieved a root mean square error (RMSE) of 86.95 and an R^2 of 0.95. In addition, SHAP analysis revealed that floor area, wall U-value, and window-to-wall ratio were the most influential variables affecting the prediction results. These results suggest that the proposed model has potential applicability for predicting annual energy consumption of office buildings based on various building characteristics.

키워드 : 에너지 소비량, 사무용 건축물, 머신러닝, 트리 기반 알고리즘

Keywords : Energy consumption, Office building, Machine learning, Tree-based algorithm

1. 서론

1.1 연구 배경

국제 에너지 기구(International Energy Agency, IEA)의 보고서에 따르면 전체 에너지 사용량 중 건물 부문 에너지 사용량은 약 40%에 이르며, 그중 냉·난방에 사용되는 에너지는 약 50%를 차지하고 있다[1]. 건물 운영 단계에서의 과도한 에너지 소비는 온실가스 배출의 주된 원인이므로, 이를 효과적으로 저감하는 것은 기후 변화 대응을 위한 과제로 부각되고 있다[2].

이에 따라, 건물 부문에서 온실가스 배출량을 감소시키는 방안으로 국내외에서는 제로에너지빌딩(Zero energy building) 도입, 그린리모델링(Green remodeling) 등 다양한 정책이 진행되고 있다. 이러한 정책의 실질적인 적용성을 높이기 위해서는 건물의 에너지 소비를 정확히 평가하고 예측할 수 있는 평가도구가 필수적이다.

현재 건물의 에너지 평가에는 주로 물리 기반 시뮬레이션 모델이 활용되고 있다. 하지만 이러한 모델들은 복잡한 계산 과정과 방대한 입력 데이터 요구로

인해 부담이 존재한다. 이러한 한계를 극복하기 위한 대안으로, 과거 데이터를 기반으로 예측이 가능한 머신러닝(Machine learning) 기법이 주목받고 있다[3]. 특히, 트리 기반 알고리즘은 예측 결과에 미치는 복잡하고 비선형적인 요인들 간의 관계를 학습하는 데 강점을 보여, 이를 통해 예측 결과의 정확도와 효율성을 높일 수 있다.

1.2 연구 목적

본 연구의 목적은 건물의 다양한 특성을 반영하여 일반화 가능한 사무용 건물 에너지 소비량 예측모델을 개발하는 것이다. 이를 위해 트리 기반 알고리즘인 Extra Trees Regressor를 활용하여 건물 특성 데이터를 기반으로 연간 에너지 소비량을 예측하는 모델을 구축하였다. 이를 통해 건물 특성 변수와 에너지 소비량 간의 관계를 해석하고 데이터 기반 예측모델의 활용 가능성을 검토하는 것을 목적으로 한다.

2. 연구 방법

2.1 데이터 수집 및 구성

본 연구에서는 사무용 건물의 연간 에너지 소비량 예측모델 개발을 위해 미국 에너지부(Department of energy, DOE)에서 제공하는 ComStock 데이터셋을 활용하였다. DOE의 건물 기후 구분 중 North Carolina

* 중앙대 대학원 석사과정

** 중앙대 대학원 박사과정

*** 중앙대 건축학부 교수, 공학박사

(Corresponding author : Department of Architectural Engineering, Chung-Ang University, gilerbert73@cau.ac.kr)

이 논문은 2026년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2023-00217322)

지역의 4A 기후대를 대상으로 데이터를 선정하였다. 해당 지역은 혼합 습윤 기후(Mixed-humid climate)에 해당하며 냉방과 난방이 모두 요구되는 기후 조건을 가지고 있어 건물 운영 환경에서 나타나는 다양한 에너지 사용 특성을 포함할 수 있다. 데이터 전처리 과정에서는 피어슨 상관계수(Pearson correlation coefficient)를 활용하여 상관계수가 0.3 이상이거나 -0.3 이하인 변수들을 최종 입력변수로 구성하였다. 이를 통해 예측에 유의미한 변수만을 선별하고 노이즈를 최소화하였다.

표1. 에너지소비량 예측모델의 입력변수 및 출력변수

Input	Stories, Climate, Floor Area [m ²], WWR, Cooling COP, Heating COP, Wall U-value [W/m ² · K], Roof U-value [W/m ² · K], Annual Total Energy Consumption [kwh/year]
Output	Annual Total Energy Consumption [kwh/year]

2.2 에너지 소비량 예측모델 개발

본 연구에서는 건물의 연간 에너지 소비량을 예측하기 위해 트리 기반 앙상블 학습 알고리즘인 Extra Trees Regressor는 다수의 결정트리(Decision tree)를 생성하여 예측 결과를 평균화하는 방식의 앙상블 모델로, 학습 과정에서 분할 기준과 분할 지점을 무작위로 선택하는 강한 랜덤성을 특징으로 한다. 이러한 특성은 모델의 분산을 줄이고 과적합을 완화하여 복잡하고 비선형적인 건물 에너지 소비 패턴을 안정적으로 학습할 수 있다.

모델은 300개의 트리를 기반으로 구성되었으며, 각 트리의 최대 깊이는 25로 설정하였다. 또한 과적합을 방지하기 위해 리프 노드의 최소 샘플 수는 20, 노드 분할을 위한 최소 샘플 수는 40으로 설정하였다. 모델의 일반화 성능을 평가하기 위해 5-Fold Cross Validation 기법을 적용하였다. 전체 55,385개의 데이터셋은 5개의 폴드로 분할되었으며, 각 반복에서 4개의 폴드를 학습 데이터로 사용하고 나머지 1개 폴드를 검증 데이터로 사용하였다. 교차 검증으로 모델의 일반화 성능을 확인한 후, 예측 성능을 극대화하기 위해 학습된 최적의 파라미터를 사용해 전체 데이터를 재학습하였다.

3. 연구 결과

3.1 에너지 소비량 예측모델 정확도

예측모델의 성능평가는 RMSE (Root Mean Square Error)와 R²를 사용하였다. RMSE는 0에 가까울수록, R²는 1에 가까울수록 성능이 우수한 것을 의미한다. 5개 폴드의 교차 검증 결과, 모델의 평균 RMSE는 86.95 그리고 R²는 0.95로 나타나 높은 정확도를 보였다.

3.2 SHAP 분석

모델의 예측 결과에 영향을 미치는 주요 요인을 파악하기 위해 SHAP (SHapley Additive exPlanations) 분석을 수행하였다. 분석 결과, 그림 1은 에너지 소비량

예측에 영향을 미치는 상위 5개의 입력변수를 나타낸다. 그 중 건물의 연면적, 벽 열관류율, 창면적비가 높은 중요도를 보이는 변수로 나타났다. 이러한 분석 결과는 모델이 물리적인 에너지 성능 요소를 중요하게 고려하고 있음을 보여준다. 그러나 동일한 연면적 구간에서도 에너지 소비량의 분포가 넓게 나타나, 연면적 외에도 건물의 외피 성능, 설비 효율, 운영 특성 등 다양한 요인 또한 영향을 미치는 것으로 판단된다.

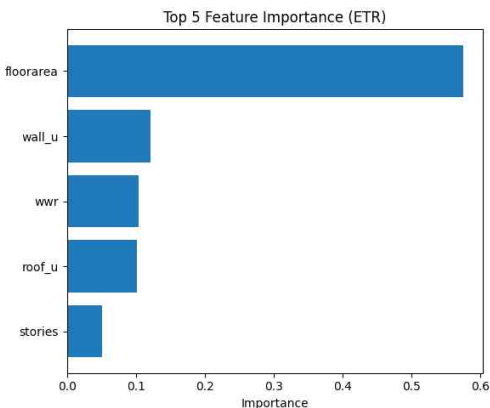


그림 1 SHAP 분석을 통한 예측모델 변수 중요도 결과

4. 결론

본 연구에서는 트리 기반 알고리즘인 Extra Trees Regressor를 활용하여 사무용 연간 에너지 소비량 예측모델을 개발하였다. SHAP 분석 결과, 건물 연면적, 벽 열관류율, 창면적비가 에너지 소비량 예측에 주요한 영향을 미치는 변수로 나타났으며, 동일한 연면적 구간에서도 에너지 소비량 분포의 폭이 넓게 나타나 다양한 건물 특성이 복합적으로 작용하는 것으로 확인되었다. 그러나 본 연구에서 개발된 예측모델은 건물 운영 과정에서 발생하는 동적 특성을 충분히 반영하지 못하는 한계가 있다. 따라서 향후 연구에서 HVAC 시스템 운영 프로파일, 재실자 밀도 변화, 내부 발열 부하와 같은 동적 요인들을 반영한 연구가 진행될 예정이다.

참고문헌

1. IEA (International Energy Agency), IEA world energy statistics and balances, 2023
2. 박동혁, 김태원, 변재운, 문진우. (2023). 태양광열-공기열원 하이브리드 히트펌프 축열조 온도 및 에너지 사용량 예측모델 개발. KIEAE Journal, 23(1), 31-36, 10.12813/kieae.2023.23.1.031
3. Mohan, R., & Pachauri, N. (2025). An ensemble model for the energy consumption prediction of residential buildings. Energy, 314, 134255. <https://doi.org/10.1016/j.energy.2024.134255>