



RESEARCH ARTICLE

건물의 에너지 벤치마크를 위한 LightGBM기반 분위 회귀 모델

Jean d'Amour Nzarigema¹ · Jack Ngarambe² · 조정훈³ · 윤근영^{4*}

¹경희대학교 건축공학과 석사과정, ²경희대학교 건축공학과 박사후연구원,
³경희대학교 건축공학과 박사과정, ⁴경희대학교 건축공학과 교수

Quantile Regression Modelling with LightGBM for Building Energy Benchmarking

Nzarigema, Jean d'Amour¹ · Ngarambe, Jack² · Zo, Chung-Hoon³ · Yun, Geun-Young^{4*}

¹Master Student, Department of Architecture, Kyung Hee University, Suwon, Korea
²Postdoctoral Researcher, Department of Architecture, Kyung Hee University, Suwon, Korea
³Ph.D Student, Department of Architecture, Kyung Hee University, Suwon, Korea
⁴Professor, Department of Architecture, Kyung Hee University, Suwon, Korea

*Corresponding author: Yun, Geun-Young, Tel: +82-31-201-3859, E-mail: ggyun@khu.ac.kr

ABSTRACT

Building energy benchmarking is a crucial step in building energy monitoring. Current benchmarking practices often face limitations rooted in the adopted methodologies. For instance, existing benchmark tools are based on the ordinary least squares regression method, which is highly sensitive to outliers and non-linearity in the dataset; this is likely to produce inaccurate benchmark scores. We propose a quantile regression benchmarking model reinforced with a gradient boosting decision tree algorithm (LightGBM). On the one hand, quantile regression analysis estimates the conditional distribution across quantiles, better capturing non-linearity in building energy behavior at various hierarchies of energy usage. On the other hand, the tree-based algorithm corrects previous errors during model training for better optimization. The said models were developed using a dataset from kindergarten school buildings in South Korea consisting of 3,090 buildings. From the developed models, it was observed that total floor area is a major determining factor in the electric energy use of kindergarten buildings in Korea, while the number of floors was also an energy driving factor in middle energy-consuming buildings. The proposed model provides accurate estimations of building energy performance, better facilitating energy comparison of a given building with its peers. The developed models provide valuable insights for energy conservation programs that might later inform building policies and for individuals wishing to reduce the energy consumption of their facilities.

주요어 : 건물에너지사용량, 에너지 벤치마크, LightGBM, 분위회귀모델, 유치원

Keywords: Building energy used, Energy benchmarking, LightGBM, Quantile regression, Kindergarten

OPEN ACCESS

Journal of KIAEBS 2022 October, 16(5): 359-373
<https://doi.org/10.22696/jkiaeb.20220031>

pISSN : 1976-6483
eISSN : 2586-0666

Received: September 16, 2022

Revised: October 16, 2022

Accepted: October 19, 2022

© 2022 Korean Institute of Architectural Sustainable Environment and Building Systems.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

서론

전 세계 에너지 소비량 중 건물에서 소비하는 에너지 사용량은 약 50%이다(Delcroix et al., 2021). 건물에서 소비되는 에너지사용에 따라 탄소배출이 높아지고 있으며, 이로 인해 지구의 온난화를 더욱 악화시킬 것으로 예상된다(Kim et al., 2017; Kim et al., 2019a; Kim et al., 2019b). 따라서 건물 에너지 절감 방안은 지구의 온난화 발생을 낮추고 나아가 경제에 큰 영향을 미치는 중요한 고려사항이다(Omer, 2008). 건물 에너지 절감 방안 중 하나인 에너지 벤치마크 프로세스는 건물의 설계 초기 단계에서 시행하는 방안으로 건물의 에너지 소비량과 주요 변수들(예: 건물의 기능, 사용 일정, 지리적 위치, 건물 크기 등)을 고려하여 에너지 소비량을 점수화하여 건물의 에너지 소비량을 효과적으로 비교할 수 있다.

에너지 벤치마크 모델은 국내외 여러 기관에서 개발되었다(Chung, 2011). 대표적인 에너지 벤치마크 모델은 미국의 환경청에서 개발한 ENERGYSTAR portfolio가 있다. ENERGYSTAR portfolio는 고객의 건물의 에너지 성능을 전국의 유사한 건물과 비교할 수 있는 모델이며 에너지 효율 리베이트를 제공한다(Azulay et al., 2009). ENERGYSTAR portfolio는 건물 에너지 성능의 목표를 설정하고 사용자가 입력한 건물의 위치, 건축 면적, 주간 평균 운영 시간, 평균 재실자 밀도, 전기 플러그 부하 밀도 및 12개월의 월간 에너지 소비량을 기준으로 맞춤형 건물 에너지 성능 목표 또는 EUI를 산출한다(Hicks and Clough, 1998). 해외 연구기관 및 단체에서도 비슷한 유형의 에너지 벤치마크 모델을 개발하였다(영국(Shahrestani et al., 2014), 중국(Chan, 2012), 싱가포르(Arjunan et al., 2022)).

기존 에너지 벤치마크 모델은 사용자에게 효과적인 건물 에너지 절감 방안을 제공하지만, 모델 개발에 있어 이론만을 근거하여 개발되었기 때문에 실제 건물에 적용하면 오차가 발생하는 것을 확인하였다. 예를 들어, 데이터 세트의 이상값에 민감하고 선형적 관련 요인으로 제한이 있는 OLS (Ordinary Least Squares, 최소제곱법) 회귀 분석을 통한 에너지 벤치마크 모델이 있다(Roche et al., 2000). 이러한 제한이 있는 OLS 회귀 기반의 에너지 벤치마크 모델의 문제점을 보완하기 위해 Yalcintas and Aytun Ozturk (2007)와 Yalcintas (2006)의 연구에서는 머신러닝 알고리즘을 이용하여 주어진 건물의 에너지 소비량이 전체 바닥 면적과 비례한다는 가정하에 주어진 단위 면적당 에너지 소비량(EUI)을 추정하는 데이터 기반 에너지 벤치마크 모델을 개발하였다. 하지만, 여전히 보완사항이 산재하여 EUI의 추정치가 정확하지 않을 수도 있다는 것을 확인하였다. Roth and Rajagopal (2018)은 건물 에너지 벤치마크 모델의 정확도를 개선하기 위해 미국의 1,000개 상업용 건물 데이터 세트를 기반으로 하는 분위 회귀 모델을 제안했다. 조건부 평균 대신 공변량을 기반으로 응답 변수의 조건부 분포를 추정한 뒤 건물 에너지 성능에 따라 주요 요인을 결정하고 이상치(Outlier)를 처리할 때 적합한 것으로 확인하였다. 그 결과, 로지스틱 선형 회귀 모델링을 사용할 때 달성되었을 때보다 표준 오차를 줄이는 것으로 확인하였다.

본 연구에서는 Roth and Rajagopal (2018)과 유사한 방법론을 사용하지만, 최적화를 위해 모델 훈련 중에 발생한 기존 오류를 수정하여 결과값을 도출하는 Tree기반 머신러닝 알고리즘인 LightGBM (Light Gradient Boosting Model)으로 강화한 분위 회귀 모델을 개발하였

다. 본 연구를 통해 개발한 분위 회귀 모델은 분위에 걸친 조건부 분포를 추정하여 다양한 에너지 소비 계층에서 유치원 건물의 에너지 소비 패턴을 추정하였으며, 모델의 비선형성적 패턴을 분석하였다.

본 논문에서는 에너지 벤치마크 모델의 프로세스에 대한 방법론을 설명하고 결과 및 토의에서는 결과를 분석한 뒤 모델의 성능을 설명하며 결론을 제시한다.

연구 방법

에너지 벤치마크 모델은 개발에 필요한 관련 에너지, 건물 특성 및 기상 데이터를 수집하고, 데이터 전처리를 통해 이상치를 제거한 뒤 건물 에너지 소비와 에너지 소비에 대한 요인 간의 패턴에 대한 통계 분석을 진행하고 분위 회귀 모델을 개발하였다.

데이터 수집

본 모델에 필요한 데이터를 수집하기 위해 정부에서 제공하는 데이터베이스를 활용하여 국내 소재의 건물 주소를 확보하였다(Public Data Portal, 2022). 각 데이터베이스에서 유치원 시설을 분류하기 위해 유치원, 보육원, 위탁시설 등으로 키워드를 사용하여 3,090개의 유치원에 속하는 리스트를 확보하였다.

3,090개 건물의 건물 특성 정보 및 월별 전기 에너지 사용량 데이터는 국토교통부에서 제공하는 공개 데이터베이스로 수집하였다(Architecture Data Private Open System, 2022). 또한, 전국적으로 분포된 유치원에 대하여 위치별 기상 조건을 확보하기 위해 기상청에서 제공하는 자동 기상 관측소(AWS)를 이용하여 기상 데이터를 수집하였다(KMA Open MET Data Portal, 2022). 3,090개의 유치원 건물과 각 건물 인근의 자동기상관측소(AWS)와 효율적으로 매칭하기 위해 두 지점의 위도와 경도 값을 기반으로 두 지리적 지점 사이의 거리를 계산하는 Haversine 함수를 사용했다(식 (1)). Figure 1은 건물 특성 정보, 월별 전기 에너지 사용량

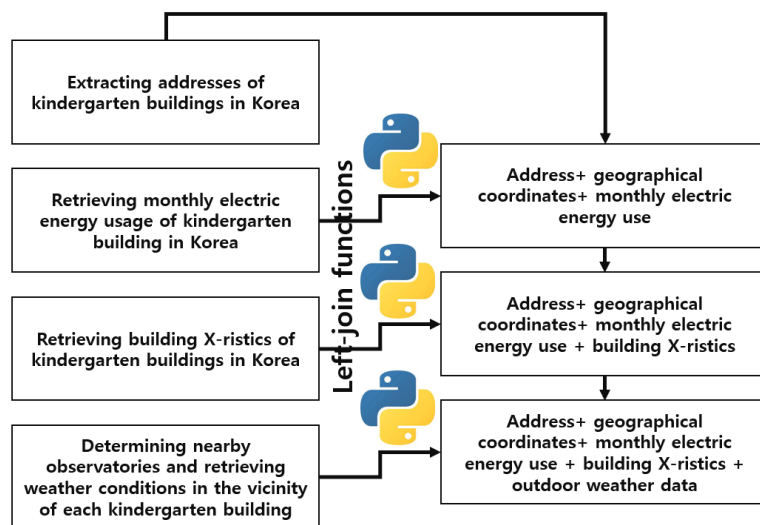


Figure 1. Data collection steps

및 기상 데이터를 취합하여 데이터베이스를 구축하는 각 단계를 나타낸다.

$$hav(\theta) = hav(\varphi_2 - \varphi_1) + (1 - hav(\varphi_1 - \varphi_2) - hav(\varphi_1 + \varphi_2)) \times hav(\lambda_2 - \lambda_1) \quad (1)$$

여기서 $hav(\theta)$ 는 haversine 함수이다. φ_1, φ_2 는 각각 포인트 1의 위도와 포인트 2의 위도이고, λ_1, λ_2 는 각각 포인트 1의 경도와 포인트 2의 경도이다.

HDD and CDDs

HDD(난방도일) 및 CDD(냉방도일)는 냉난방을 위하여 기상 조건과 관련된 에너지 소비량을 추정할 때 사용되는 계수이다. 냉난방도일은 실내의 열쾌적 환경을 제공하기 위해 건물이 필요로 하는 에너지를 나타내며 지정된 균형점 온도를 기반으로 계산된다. 균형점 온도는 국가마다 다르며, 국내의 경우 CDD는 24°C, HDD는 18°C를 주로 사용하였으며, 본 연구에서도 동일하게 적용하였다(Choi, 2005). 식 (2)와 식 (3)은 CDD와 HDD의 산출식이다.

$$CDD = \sum_{i=1}^N (T_i - T_b) \text{ (for } T_i \geq T_b) \quad (2)$$

$$HDD = \sum_{i=1}^N (T_i - T_b) \text{ (for } T_i \leq T_b) \quad (3)$$

여기서 T_i 는 i 일의 평균기온, T_b 는 기준기온, N 은 각각 일일 평균기온이 기준기온보다 높거나 낮을 때의 일수이다.

냉난방 에너지 사용량 도출

건물의 총 전기 에너지 소비량에서 냉난방 에너지 소비량을 분류하는 다양한 방법이 있다. 대표적인 방법으로는 건물에서 사용되는 전자 기기들의 세부 정보들을 파악하고 각 기기의 에너지 사용량을 기반으로 산출하는 추정 알고리즘을 기반으로 한다(Field et al., 1997; Wang et al., 2012). 본 연구에서 개발한 모델의 경우 건물에서 사용되는 전자 기기들에 대한 정보가 없다는 점을 감안하여, 냉난방을 목적으로 사용되는 에너지 소비량을 추정하기 위해 단기 측정법(STM, Short-Term Measurement Method)을 사용했다. 5월과 10월 같이 봄과 가을과 같은 중간기의 전기 에너지 사용량 중 최소 에너지 사용량을 기준값으로 총 월별 전기 에너지 사용량과 기준값 간의 차이를 산출하여 3,090개 건물 각각에 냉방과 난방 에너지 소비량을 선정하였다(Robison, 1992).

모델 개발

분위 회귀 모델

다양한 예측 변수에 걸쳐 응답 변수의 조건부 평균값을 추정하는 선형 회귀 분석과 달리 분위 회귀 분석은 반응 변수의 조건부 분위값을 추정한다(Rodriguez and Yao, 2017). 분위 회귀 모델은 선형 회귀 분석의 조건이 충족되지 않을 때 적합한 선형 회귀 분석의 확장이며 식 (4)에 따라 산출된다. 여기서, 베타 계수는 선형 회귀 분석과 달리 일정하지 않으며, 분위에 따라 달라지는 함수이다. 각 분위의 베타 계수는 중위수 절대 편차를 최소화함으로써 찾을 수 있다. 본 연구에서는 99개의 분위기를 대표하는 99개 모델을 개발하였다.

$$q_{\tau}(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \dots + \beta_z(\tau)x_{iz}, i = 1, \dots, n \quad (4)$$

여기서 y_i 는 응답 변수이고 β_0 은 특정 분위 β 에서의 절편이며, x 는 베타 계수, x 는 예측 변수, z 는 예측 변수의 수, n 은 데이터 포인트의 수이다.

LightGBM

LightGBM은 빠른 처리, 높은 정확도 및 빠른 훈련 속도로 대용량 데이터를 처리에 적합하며, 메모리를 효율적으로 활용한다(AI Daoud, 2019). 또한, 모델의 성능을 극대화하기 위해 하이퍼 파라미터의 최적화로 모델 성능을 높일 수 있게 최상의 모델 매개변수를 검색하는 프로세스를 제공한다. 본 연구로 개발한 모델에서는 하이퍼 파라미터의 랜덤 서치(Random Search) 기능을 사용하고 하이퍼 파라미터 최적화 프로세스를 자동화시킨 뒤 개발된 99개 분위 회귀 모델에 대하여 각각의 하이퍼 파라미터를 결정 후 적용하였다.

모델 성능 평가

본 연구에서는 결정계수(R^2)와 수정결정계수(Adjusted R^2)를 사용하여 모델 성능을 평가했다. 결정계수는 종속변수에 대한 독립변수들의 설명력을 알고자 할 때 사용되며, 수정결정계수는 결정계수를 표본 수와 독립변수의 개수를 기반으로 보정되는 계수이다. 본 연구에서는 개발된 99개 모델 각각에 대해 두 가지 계수를 모두 계산하여 평가했다.

모델 입력, 출력 및 벤치마크 로직

본 연구의 에너지 벤치마크 모델은 건물의 특성 정보와 기상 조건을 사용하여 분위 점수를 산출한다. 본 에너지 벤치마크 모델은 10개의 변수(대지면적, 연면적, 준공연도, 층수, 용적률, 외기 온도, 최소 외기 온도, 최대 외기 온도, 난방도일, 냉방도일)를 입력하면, 99개 분위에 걸쳐 산출된 건물의 전기 에너지 소비를 예측하며, 사용자가 입력한 1년간의 월별 실제 전기 에너지 사용 데이터를 토대로 해당 건물 유형의 가장 가까운 분위기를 결정하기 위해 모델의 99개 추정치와 반복적으로 비교하고 실제 에너지 사용에 가장 가까운 에너지 추정치를 가진 분

위를 도출한다. 이때, 각 입력 변수인 기상 조건 및 건물 특성 정보는 통계적으로 유의한 상관 관계를 가진다. 특히 건물의 크기(예: 건축 면적, 층수) 및 준공연도와 같은 변수가 건물 에너지 소비량과 높은 상관관계를 가진다는 기존 연구를 참고하여 채택하였다(Cheng and Steemers, 2011; Aksoezen et al., 2015; Kotsila and Polychronidou, 2021).

결과 및 토의

데이터 설명

Figure 2는 국내 3,090개 유치원 건물에서 수집한 월별 전기 에너지 사용량 분포를 보여주고 있다. Figure 2에서 볼 수 있듯이 본 모델의 데이터 세트의 비정규성을 보여주고 있으며, 에너지 소비 패턴을 모델링하기 위한 분위 회귀의 사용을 지원하는 양으로 치우친 대칭을 채택했다. 에너지 사용량의 데이터 세트는 20,000 kWh에서 다수의 건물이 분포되어 있으며, 데이터 세트의 평균(M)은 34,181 kWh, 표준 편차(SD)는 24,551, 중앙값은 34,181 kWh로 확인되었다. 이를 토대로 OLS 회귀 분석에 대하여 적합성을 검토해본 결과 평균과 중앙값 간에 큰 차이를 보였으며, 데이터가 균일하게 분포되지 않은 것으로 확인되어 본 모델에는 적합하지 않은 것으로 확인되었다.

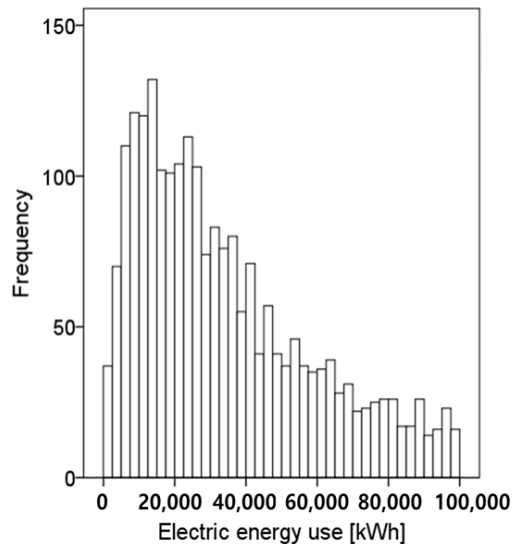


Figure 2. Distribution of monthly electric energy consumption of kindergarten school buildings in South Korea

전기 에너지 사용의 주요 요인

Figure 3은 일부 건물의 특성, 날씨 데이터 및 유치원의 전기 에너지 사용량 간의 관계를 보여준다. 데이터 세트의 이상치를 줄이고 에너지 사용과 일부 요인 간의 관계를 파악하기 위해 변수를 클러스터로 그룹화하고 해당 그룹의 에너지 변화를 평가했다. 예를 들어, 층수가 증가함에 따라 전기 에너지 사용량이 증가한다는 것을 발견했다(NoF, 건물의 층수). 예를 들어,

클러스터 1의 건물(1개 층에서 4개 층 사이)과 클러스터 7의 건물(12개 층 이상)의 평균 에너지 사용량 차이는 약 3,537,498 kWh였다. 건축 면적을 고려할 때도 비슷한 경향이 나타났다. 예상대로 건축 면적이 클수록 더 많은 에너지 전기 에너지가 소비된다. 또한, 클러스터 1의 건물(5,000 m² 이하의 건축 면적)의 경우 평균적으로 클러스터 4의 건물(16,300 m² 이상의 건축 면적)보다 약 7,012,5654 kWh 더 많이 소비한다. HDD 및 CDD도 건물 에너지 사용과 양의 상관관계를 확인하였다. 높은 HDD 및 CDD 클러스터는 대부분 높은 건물 에너지 사용과 관련이 있다. 예를 들어, CDD가 1,213°C-hour 이하인 클러스터 1은 클러스터 6(HDD 범위가 4,854~6,067°C-hour)보다 377,237 kWh 적은 평균 전기 에너지 사용에 해당한다. HDD를 고려할 때도 유사한 경향이 관찰되었다 (Table 1).

Table 1. Range of values in each cluster

Variable	Cluster									
	1	2	3	4	5	6	7	8	9	10
1	100	7692.5	15284.8	≥						
	-	-	-	22877.2						
	7692.4	15284.7	22877.1							
2	100	14133.2	28166.2	42199.3	≥					
	-	-	-	-	56232.3					
	14133.1	28166.1	42199.2	56232.2						
3	1.5	601.7	1201.8	≥						
	-	-	-	1801.9						
	601.6	1201.7	1801.8							
4	1	5	9	13	18	22	≥			
	-	-	-	-	-	-	25.1			
	4	8	12	17	21	25				
5	1	12	23	34	45	≥				
	-	-	-	-	-	55.1				
	11	22	33	44	55					
6	7.3	7.4	9.4	11.4	13.4	≥	15.5			
	≤	-	-	-	-	-				
		9.4	11.4	13.4	15.4					
7	-17.8	-17.9	-15.9	-13.9	-11.9	-9.9	-7.9	-5.9	-3.9	≥
	≤	-	-	-	-	-	-	-	-	-1.9
		-16	-14	-12	-10	-8	-6	-4	-2	
8	27	27.1	29.1	31.1	33.1	35.1	≥	37.2		
	≤	-	-	-	-	-				
		29.1	31.1	33.1	35.1	37.1				
9	1213.4	1213.5	2427.1	3640.6	>					
	≤	-	-	-	4854.1					
		2427	3640.5	4854						
10	23758.6	23758.7	37625.4	51492.1	65358.8	>				
	≤	-	-	-	-	79225.4				
		37625.3	51492	65358.7	79225.3					

1 = building area (m²) , 2 = total floor area (m²), 3 = FAR, 4 = NoF, 5 = building age (years), 6 = outdoor temperature, 7 = Minimum monthly temperature, 8 = Maximum monthly temperature, 9 = CDD (°C-hours), 10 = HDD (°C-hours)

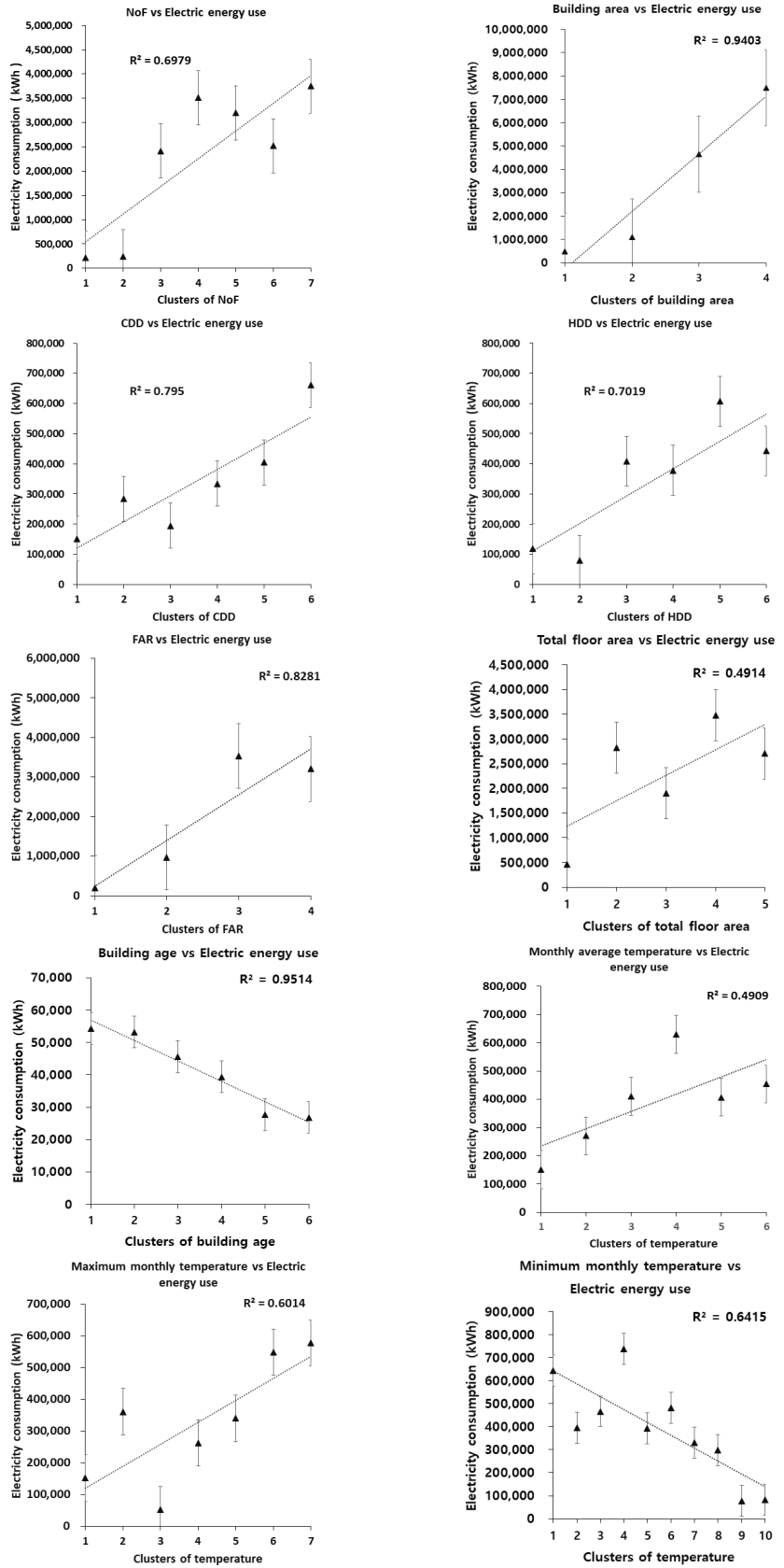


Figure 3. Relationship between various factors and monthly electric energy use of kindergarten buildings

LightGBM 기반 분위 회귀 모델의 성능

앞서 간략히 논의한 바와 같이 개발된 모델은 99개의 분위기를 기반으로 했다. 성능 관련 지표, 특히 결정계수(R^2) 및 수정결정계수(Adjusted R^2)는 모델이 특정 분위 내에서 에너지 사용을 얼마나 정확하게 예측했는지 평가하기 위해 각 분위에 대해 산출했다.

아래 Table 2는 다양한 분위에서 일부 모델의 성능을 보여준다. 통계적 유의성은 P 값 < 0.05에서 결정되었다. 예를 들어, 10번째 분위에서의 결정계수 0.18은 10번째 백분위에 있는 건물의 전기 에너지 소비 변동의 18%가 고려된 설명 변수의 변화로 설명되며, 80번째 분위에서의 결정계수 0.58은 80번째 백분위에서 건물의 전기 에너지 사용 변동의 50%가 고려된 회귀 변수로 설명된다. 각 분위에 걸친 결정계수와 수정결정계수는 입력 변수에 따라 분위 회귀 모델 성능에 미치는 결과를 보여준다.

Table 2. Performance of the developed models at example quantile

Quantile	R^2	Adjusted R^2	P value
10th	0.1896	0.187	< 0.05
20th	0.3281	0.3259	< 0.05
30th	0.3549	0.3528	< 0.05
40th	0.4365	0.4346	< 0.05
50th	0.4874	0.4857	< 0.05
60th	0.5571	0.5557	< 0.05
70th	0.5246	0.523	< 0.05
80th	0.5804	0.579	< 0.05
90th	0.3238	0.3216	< 0.05

전기 에너지 소비량에 대한 각 입력 변수의 영향을 정량화하는 매개변수 추정치는 통계적 유의성과 함께 25분위, 50분위 및 75분위에 대해 Table 3에 나와 있다. 모든 입력 변수의 영향은 통계적으로 유의하지만($P < 0.05$), 영향의 크기와 방향은 변수와 분위에 따라 다르다. 예를 들어, 분위 회귀 모델의 건축면적 변수에 대한 계수는 25분위에서는 11.08로 나타났으나, 75분위에서는 115.10으로 증가하였다. 이와 같은 분석 결과는 에너지 소비량과 건축 면적이 관련되어 있다는 것을 나타낸다. 또한, 이와 유사하게, 건물의 준공연도는 각 분위에 걸쳐 감소하는 것으로 분석되었다. 즉, 분위 회귀 모델의 준공연도 계수는 25분위에서 -65.63인 반면, 75분위에서는 -1263.30로 감소했다. 또한, 에너지 소비량이 큰 건물의 경우 건물의 준공연도에 대한 영향이 상대적으로 낮은 것으로 확인되었다.

Table 3의 양(+)과 음(-)의 기호는 전기 에너지 사용량의 증가 또는 감소에 따른 각 변수 간의 관계의 방향을 나타낸다. 그리고, 변동계수(coefficient)는 입력 변수의 값이 단위 변화당 전기 에너지 소비량의 증가치를 나타낸다.

Table 3. Statistical significance of the effect of each variable on building energy use

Variable	Quantile								
	25th			50th			75th		
	Parameter estimates								
	coefficient	t-statistic	P value	coefficient	t-statistic	P value	coefficient	t-statistics	P value
1	-15459.58	-0.88	< 0.05	15153.69	0.39	< 0.05	-23627.31	-0.19	0.85
2	11.08	15.49	< 0.05	8.93	5.72	< 0.05	115.1	22.4	< 0.05
3	47.54	188.35	< 0.05	95.51	173.31	< 0.05	175.61	96.85	< 0.05
4	31.79	63.1	< 0.05	15.87	14.42	< 0.05	242.32	66.96	< 0.05
5	5281.45	22.99	< 0.05	24026.36	47.91	< 0.05	52142.35	31.6	< 0.05
6	-65.63	-2.03	< 0.05	-273.98	-1.84	< 0.05	-1263.3	-2.58	< 0.05
7	-2781.91	-2.03	< 0.05	2920.64	0.97	< 0.05	125.71	0.01	< 0.05
8	1117.46	1.46	< 0.05	761.35	0.46	< 0.05	1409.52	0.26	< 0.05
9	621.59	0.7	< 0.05	-5980.8	-3.09	< 0.05	-3732.46	-0.59	< 0.05
10	-0.51	-0.56	< 0.05	1.59	0.79	< 0.05	-7.36	-1.11	< 0.05
11	0.26	1.41	< 0.05	1.44	3.54	< 0.05	1.03	0.77	< 0.05

1=intercept, 2=building area (m²), 3= total floor area (m²), 4=floor area ratio, 5=NoF, 6=building age (years), 7=outdoor temperature (°C), 8=minimum monthly temperature (°C), 9=maximum monthly temperature (°C), 10=CDD (°C-hours), 11= HDD (°C-hours)

각 분위 모델별 변수의 영향도 변화

Figure 4는 분위 회귀 모델에 포함된 변수가 각 분위에서 전기 에너지 소비에 대한 영향도를 보여준다. OLS 회귀 분석 모델과 건물 에너지 벤치마크를 위한 분위 회귀 모델의 비교를 위해 OLS 회귀 분석을 사용할 때 각 분위에 매개변수 추정치(빨간색으로 표기된 부분)를 포함하였다.

Figure 4에서 볼 수 있듯이 OLS 회귀 분석에 의한 매개변수 추정치는 분위에 걸쳐 일정하지만, 분위 회귀 모델에 의한 추정치는 각 변수의 영향 크기를 나타내는 분위의 함수로 다양하다. 예를 들어, 건축 면적의 영향은 낮은 분위(60번째 분위 미만)에서 상대적으로 낮지만 60번째 분위보다 높은 분위에 대해 상당히 증가한다는 것이 관찰되었다. 이는 건축 면적이 에너지 소비가 많은 건물에 미치는 영향이 에너지 소비가 적은 건물에 미치는 영향보다 훨씬 더 크다는 것을 나타낸다. CDD의 경우 건축 면적과 유사한 추세를 띠지만, HDD의 경우 그 반대로 관찰되었다. OLS 회귀 분석에서는 가능하지 않지만, 잠재적인 보존 조치와 관련된 중요한 세부 정보를 제공한다. 예를 들어, 상위 분위에서 관찰된 CDD의 높은 매개변수 추정치는 국내의 유치원에 효율이 높은 냉방 설비가 필요한 것으로 해석될 수 있다. 분위 회귀 모델의 영향 플롯 평가는 스펙트럼의 극단에 있는 건물 에너지 요인에 대한 척도를 제공한다.

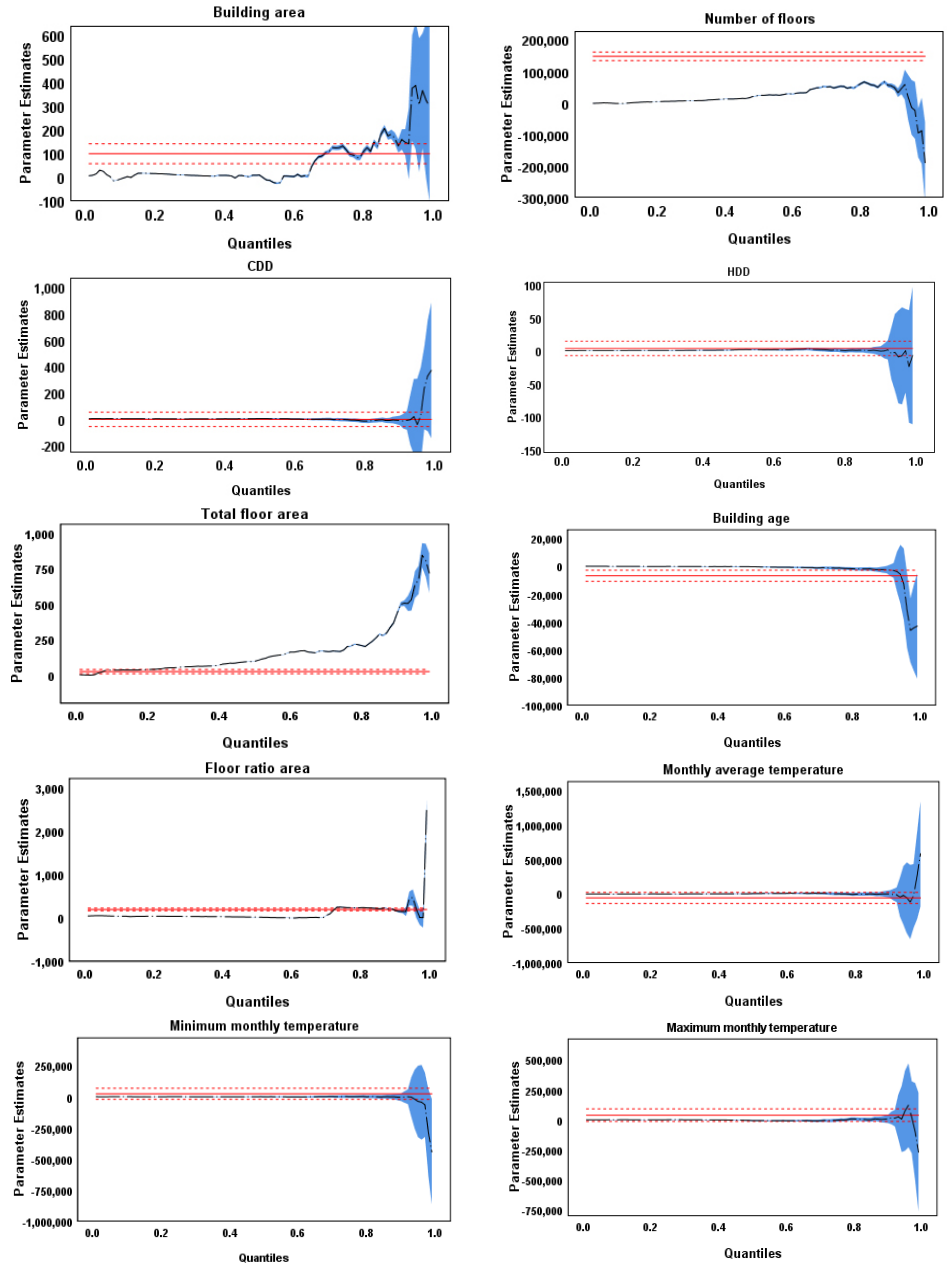


Figure 4. Influence plots showing the variations of parameters estimates of example variables across quantile

에너지 벤치마크 프로세스

본 연구에서 개발한 모델은 건물의 특성 정보와 기상 정보가 혼합된 10개의 입력값을 받아 분위 점수와 벤치마크 점수를 산출한다. 본 모델은 분석의 기반이 되는 99개의 분위기를 반영하는 99개의 결과를 다음의 프로세스에 따라 산출된다.

- (1) Figure 3에 표시되는 10개의 변수를 입력
- (2) 입력값을 이용하여 99개 분위에 걸친 건물의 전기 에너지 소비량을 추정
- (3) 에너지 벤치마크 건물에 대한 과거 1년의 전기 에너지 사용량 입력

(4) 실제 에너지 사용량을 분위 회귀 모델로 산출된 에너지 소비 추정치와 반복 비교 후 오차가 가장 낮은 분위의 점수를 산출

본 모델의 최종 목표는 사용자가 입력한 건물의 에너지 소비 분위기를 결정하는 것이다. 분위가 높을수록 건물에서 더 많은 에너지를 소비하는 것을 뜻하며, 이러한 논리를 바탕으로 해당 건물의 에너지 성능을 다른 건물과 비교하여 추론할 수 있다. Figure 5는 모델 개발에 사용된 특정 입력과 벤치마크 로직을 보여준다.

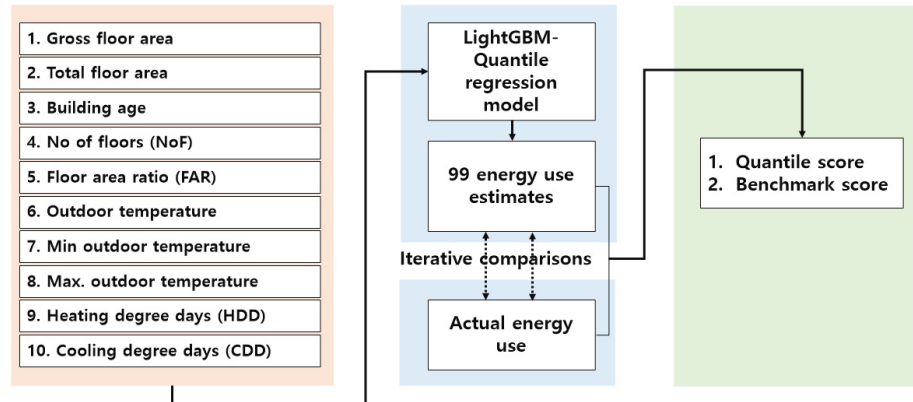


Figure 5. Model inputs, outputs, and benchmarking logic

Figure 6은 모델 추정치에서 생성된 단일 예시 건물에 대한 누적 분포 함수를 보여준다. 개발된 모델에 활용된 예시 건물의 10가지 입력 변수는 건축 면적(379.43 m²), 연면적(1828.48 m²), 준공연도(2019), FAR(445.18), NoF(4개 층), 월 평균 기온(13.30°C), 최소 월간 온도(-9.88°C), 최대 월간 온도(35.89°C), 냉방도일(4,822.94°C-hour) 및 난방도일(59,918.87°C-hour)이다.

누적 분포 곡선에서 실제 연간 전기 에너지 소비를 알고 있다는 점을 감안할 때 예제 건물의 분위 점수를 결정할 수 있다. 설명을 위해 예제 건물의 실제 연간 전기 에너지 소비량(예: 400,000 kWh)을 고려하였다. X축의 400,000 kWh 지점을 y축의 해당 값과 정렬하면(빨간색 선 참조) 예제 건물은 에너지 소비량의 68번째 분위에 위치한다.

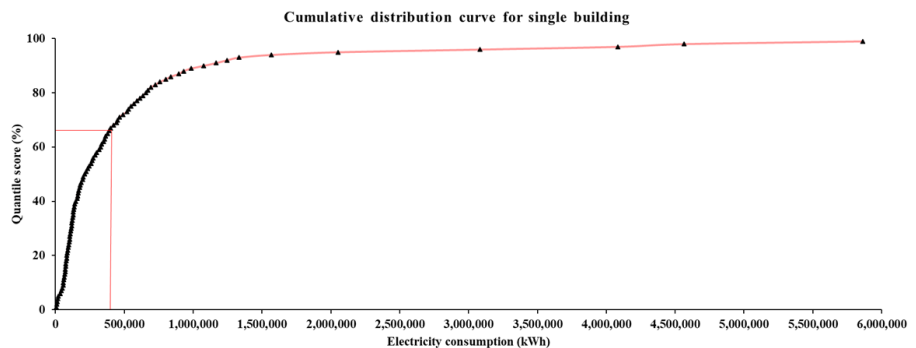


Figure 6. The distribution of building electrical energy consumption for building sample selected from the Kindergarten dataset.

결론

본 연구에서는 국내 유치원 건물에 대하여 건물 에너지 소비 계층 전반에 걸쳐 에너지 소비를 추정하는 LightGBM을 이용한 분위 회귀 분석 기반의 에너지 벤치마크 모델을 개발하였다. 본 모델은 데이터 세트의 이상치 및 비선형성에 대해 견고하여 기존 OLS 회귀 기반의 에너지 벤치마크 모델보다 좋은 대안이 될 것으로 판단한다. 또한, 모델에 의해 생성된 다양한 영향 플롯은 건물 에너지 소비 스펙트럼의 극단에 있는 주요한 요인에 대한 고려사항이 될 수 있으며, 에너지 소비량이 높은 건물의 경우 에너지 절약 조치를 시행할 때 참고 요인으로 활용될 수 있을 것으로 판단된다.

본 연구의 한계점

현재 연구의 주요 한계는 건물의 전기 에너지 소비량에서 냉난방 에너지 소비량만 고려한다는 것이다. 유치원 건물에 대한 가스 에너지 사용량 데이터베이스가 전기 에너지 사용량 데이터베이스에 비해 상대적으로 작았기 때문에 전기 및 가스 에너지를 모두 고려하게 되면 데이터베이스의 크기가 줄어들고 이로 인해 에너지 벤치마크 모델의 성능에 영향을 미칠 것으로 판단된다.

향후 전기 및 가스 에너지를 모두 고려한 에너지 벤치마크 모델에 관한 연구와 본 연구에서 제안한 에너지 벤치마크 모델과 다른 에너지 벤치마크 모델 간의 성능 비교에 관한 연구를 진행할 예정이다.

Abbreviations

EUI	: 단위 면적당 에너지 소비량
OLS	: 최소제곱법
STM	: 단기 측정법
FAR	: 연면적
NoF	: 층수
CDD	: 냉방도일
HDD	: 난방도일
Coefficient	: 변동계수

후기

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP) 연구비 지원을 받아 수행한 연구과제입니다(No.20202020800360).

References

1. Aksoezen, M., Daniel, M., Hassler, U., Kohler, N. (2015). Building age as an indicator for energy consumption. *Energy and Buildings*, 87, 74-86.
2. Al Daoud, E. (2019). Comparison between XGBoost, LightGBM and CatBoost using a home credit dataset. *International Journal of Computer and Information Engineering*, 13(1), 6-10.
3. Arjunan, P., Poolla, K., Miller, C. (2022). BEEM: Data-driven building energy benchmarking for Singapore. *Energy and Buildings*, 260, 111869.
4. Chan, W. (2012). Energy benchmarking in support of low carbon hotels: Developments, challenges, and approaches in China. *International Journal of Hospitality Management*, 31(4), 1130-1142.
5. Cheng, V., Steemers, K. (2011). Modelling domestic energy consumption at district scale: A tool to support national and local energy policies. *Environmental Modelling & Software*, 26(10), 1186-1198.
6. Choi, Y.G. (2005). Temporal and spatial variability of heating and cooling degree-days in South Korea, 1973-2002. *Journal of the Korean Geographical Society*, 40(5), 584-593.
7. Chung, W. (2011). Review of building energy-use performance benchmarking methodologies. *Applied Energy*, 88(5), 1470-1479.
8. Field, J., Soper, J., Jones, P., Bordass, W., Grigg, P. (1997). Energy performance of occupied non-domestic buildings: Assessment by analysing end-use energy consumptions. *Building Services Engineering Research and Technology*, 18(1), 39-46.
9. Kim, D.W., Kim, Y.M., Lee, S.E. (2019a). Development of an energy benchmarking database based on cost-effective energy performance indicators: Case study on public buildings in South Korea. *Energy and Buildings*, 191, 104-116.
10. Kotsila, D., Polychronidou, P. (2021). Determinants of household electricity consumption in Greece: a statistical analysis. *Journal of Innovation and Entrepreneurship*, 10(1), 1-20.
11. Omer, A.M. (2008). Energy, environment and sustainable development. *Renewable and Sustainable Energy Reviews*, 12(9), 2265-2300.
12. Roche, L., Dewey, E., Littlefair, P. (2000). Occupant reactions to daylight in offices. *International Journal of Lighting Research and Technology*, 32(3), 119-126.
13. Roth, J., Rajagopal, R. (2018). Benchmarking building energy efficiency using quantile regression. *Energy*, 152, 866-876.
14. Shahrestani, M., Yao, R., Cook, G.K. (2014). A review of existing building benchmarks and the development of a set of reference office buildings for England and Wales. *Intelligent Buildings International*, 6(1), 41-64.
15. Wang, S., Yan, C., Xiao, F. (2012). Quantitative energy performance assessment methods for existing buildings. *Energy and Buildings*, 55, 873-888.
16. Yalcintas, M. (2006). An energy benchmarking model based on artificial neural network method with a case example for tropical climates. *International Journal of Energy Research*, 30(14), 1158-1174.

17. Yalcintas, M., Aytun Ozturk, U. (2007). An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database. *International Journal of Energy Research*, 31(4), 412-421.
18. Delcroix, B., Sansregret, S., Martin, G.L., Daoud, A. (2021, November). Quantile regression using gradient boosted decision trees for daily residential energy load disaggregation. In *Journal of Physics: Conference Series*, 2069(1), 012107, IOP Publishing.
19. Hicks, T., Clough, D. (1998). The ENERGY STAR building label: Building performance through benchmarking and recognition. In *Proceedings of the ACEEE 1998 Summer Study on Energy Efficiency in Buildings*, 4, 205-210.
20. Kim, D.W., Kim, Y.M., Lee, S.H., Park, W.Y., Bok, Y.J., Ha, S.K., Lee, S.E. (2017, August). Development of reference building energy models for South Korea. In *Proceedings of the 15th IBPSA Conference*, San Francisco, CA, USA, 7-9.
21. Rodriguez, R.N., Yao, Y. (2017, April). Five things you should know about quantile regression. In *Proceedings of the SAS Global Forum 2017 Conference*, Orlando, FL, USA, 2-5.
22. Azulay, G., Barclay, N.D., Tolkin, B. (2009). ENERGY STAR® Portfolio Manager and Utility Benchmarking Programs: Effectiveness as a Conduit to Utility Energy Efficiency Programs.
23. Kim, T., Kang, B., Kim, H., Park, C., Hong, W.H. (2019b). The study on the Energy Consumption of middle school facilities in Daegu, Korea. *Energy Reports*, 5, 993-1000.
24. Robison, D. (1992). Pacific power: the use of short-term measurements to decompose commercial billing data into primary end uses. In *ACEEE 1992 Summer Study on Energy Efficiency in Buildings*, vol. 3. ACEEE Washington, DC.
25. Public Data Portal. (2022). Available at: <https://open.eais.go.kr/main/main.do>.
26. Architecture Data Private Open System. (2022). Available at: <https://www.data.go.kr>.
27. KMA Open MET Data Portal. (2022). Available at: <https://data.kma.go.kr/cmnm/main.do>.