

과거 데이터 분석을 통한 원룸의 가스사용량 예측

박준영, 송광호, 정용안, 한대진, 나웅수*
공주대학교

jypark@bitlab.re.kr, skh990427@naver.com, dyddks0115@naver.com, djhan8733@gmail.com

*wsna@kongju.ac.kr

Predicting gas consumption in a studio apartment based on past data analysis

JunYyoung Park, Kwangho Song, Yongan Cheong, Deajin Han, Woongsoo Na*
Kongju National University

요약

본 연구에서는 사용자의 이전 가스 사용량 패턴을 기반으로 미래의 가스 소비를 예측하기 위해 여러 회귀 모델을 사용했습니다. 제안하는 예측모델은 지역별 도시가스 공급업체가 제공하는 지역 및 거주 유형별 가스 사용 데이터를 기반으로 하여 온도, 공휴일등 여러 요인을 통합함. 예측하고자 하는 달의 예상 평균 기온과 이전 가스 사용량을 입력함으로써 제안하는 모델은 미래의 가스 사용량을 예측할 수 있음.

I. 서론

제안한 연구에서는 양산시, 울산시 지역 도시가스 공급업체에서 제공하는 과거 가스 사용 데이터를 기반으로, 평균 온도, 특정 온도 구간의 월별 빈도수, 공휴일 수, 그리고 평균 온도의 표준편차를 고려하여 거주 유형에 따라 분류하고자 했음. 이후 원하는 시점의 예상 평균 기온과 과거 가스 사용량을 입력하여 회귀 모델을 활용하여 가스 사용량을 예측했음.

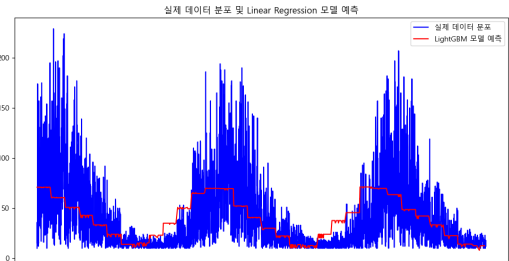
II. 연구 방법

본 연구에서는 도시가스 공급업체에서 제공하는 년도 별 고객 과금 정보 csv 파일에서 불필요한 비식별화 되어있는 정보들을 삭제하고, 여러 항목들을 추가시키며 LinearRegression 모델과 LGBM 모델을 통해 회귀 분석을 진행했음.

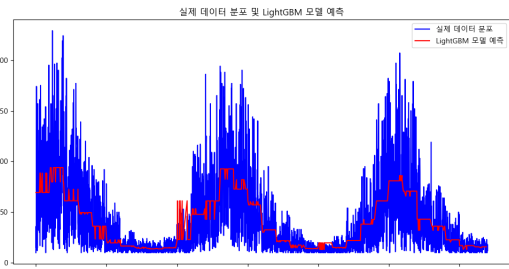
본 연구에서 제안한 모델은 두가지로 LinearRegression 을 사용한 모델과 LGBM(Light Gradient Boosting Machine)을 사용한 모델이 있음. LinearRegression 은 최소 제곱법을 사용하여 독립 변수와 종속 변수의 선형관계를 학습하는 모델임. LinearRegression 의 주요 특징은 단순하고 이해하기 쉬운 모델이며, 다중 변수를 다룰 수 있으며 각 입력 변수들의 영향력을 확인할 수 있음. LGBM 은 경사 부스팅(Gradient Boosting) 기반의 트리 앙상블 모델임. LGBM 의 주요 특징은 트리 기반 모델 중 하나로서, 기존 트리 모델보다 더 빠른 학습 속도를 가진다는 것임.

III. 결론

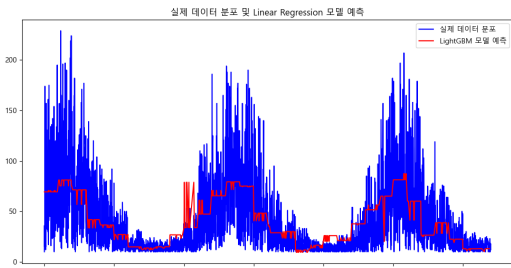
본 연구에서 데이터 파일에 처음 항목을 추가시켰을 때에는 단순하게 해당하는 월별 평균 온도만 추가시켰음. 평균 온도만 추가하고 LinearRegression 모델을 사용해서 분석을 진행했을 때는 MSE 값이 870.5 가 나왔음. 그리고 LGBM 모델을 사용해서 분석을 진행했을 때는 MSE 값이 704.3 이 나왔음. 그림 1, 2 는 평균 온도만 추가했을 때의 LinearRegression 모델과 LGBM 모델을 사용하여 회귀 분석을 진행한 그림임.



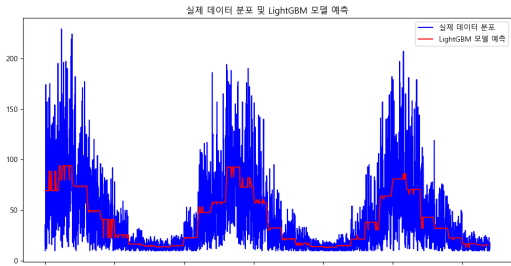
<그림 1. LinearRegression 을 사용한 모델>



<그림 2. LGBM 을 사용한 모델>



<그림 3. LinearRegression을 사용한 모델>



<그림 4. LGBM을 사용한 모델>

그림 1 과 그림 2 를 확인했을 때 원본 데이터의 모양을 어느정도 비슷하게 따라가는 모양을 보이지만 중간중간 값이 튀는 것을 확인할 수 있음. 회귀 분석에 사용되는 데이터가 너무 단순하다고 판단하여 분석에 사용되는 데이터 파일에 평균 온도와 특정 온도 구간의 월별 빈도수, 공휴일 수, 평균 온도의 표준편차를 추가했음. LinearRegression 모델을 사용해서 분석을 진행했을 때는 MSE 값이 755.3 이 나왔음. 그리고 LGBM 모델을 사용해서 분석을 진행했을 때는 MSE 값이 679 가 나왔음. 그림 3, 4 는 여러 항목을 추가했을 때 LinearRegression 모델과 LGBM 모델을 사용하여 회귀 분석을 진행한 그림임.

여러가지 항목을 추가하고 분석을 진행했을 때 MSE 값은 감소하여 정확도가 올라간 것은 볼 수 있지만 감소한 양이 너무 적어 뛰어난 성능 향상이라고 볼 수 없음. 이렇게 적은 성능 향상이 나온 이유를 살펴보았을 때, 데이터 전처리를 진행할 때 평균온도와 특정 온도 구간의 월별 빈도수를 세부 지역별로 나누지 않은 것이 원인으로 생각됨. 좀 더 세부적인 데이터를 통해 데이터 전처리를 진행한다면 더 나은 성능 향상이 있을 것이라 기대됨.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업 (IITP-2023-RS-2022-00156353) 및 2023 년도 교육부의 재원으로 중점연구소 지원사업 (2019R1A6A1A03032988)의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] 조현아, 김도연, 유다솜, 송민석, 최남승, 윤찬웅, and 박종필, "머신러닝을 활용한 가정용 도시가스 사용량 예측," 대한산업공학회 추계학술대회 논문집, pp. 2561-2585, 개최지, 개최날짜.