

XGBoost 회귀를 활용한 편의점 계약전력 예측 모델의 최적화에 대한 연구

김상민* · 박찬권** · 이지은***

A Study on the Optimization of a Contracted Power Prediction Model for Convenience Store using XGBoost Regression

Sang Min Kim* · Chankwon Park** · Ji-Eun Lee***

■ Abstract ■

This study proposes a model for predicting contracted power using electric power data collected in real time from convenience stores nationwide. By optimizing the prediction model using machine learning, it will be possible to predict the contracted power required to renew the contract of the existing convenience store. Contracted power is predicted through the XGBoost regression model. For the learning of XGBoost model, the electric power data collected for 16 months through a real-time monitoring system for convenience stores nationwide were used. The hyperparameters of the XGBoost model were tuned using the GridsearchCV, and the main features of the prediction model were identified using the `xgb.importance` function. In addition, it was also confirmed whether the preprocessing method of missing values and outliers affects the prediction of reduced power. As a result of hyperparameter tuning, an optimal model with improved predictive performance was obtained. It was found that the features of `power.2020.09`, `power.2021.02`, `area`, and `operating time` had an effect on the prediction of contracted power. As a result of the analysis, it was found that the preprocessing policy of missing values and outliers did not affect the prediction result. The proposed XGBoost regression model showed high predictive performance for contract power. Even if the preprocessing method for missing values and outliers was changed, there was no significant difference in the prediction results through hyperparameters tuning.

Keyword : XGBoost Regression, Contract Power, Prediction Model, Convenience Store

Submitted : August 20, 2022

1st Revision : August 25, 2022

Accepted : August 29, 2022

* (주)임팩스 기술연구소 수석연구원, 제1저자

** 한양사이버대학교 생산물류유통학과 교수, 공동저자

*** 한양사이버대학교 경영정보·AI비즈니스학과 교수, 교신저자

1. 서 론

2020년 10월 28일 ‘2050 탄소중립’ 선언과 12월 10일 ‘2050 탄소중립 비전’ 선포를 계기로 국가 전반의 녹색 전환을 위한 정책·사회·기술 혁신 방향으로 에너지 효율의 혁신적인 향상이 대두되었다. 이에 에너지 사용의 효율을 극대화하기 위한 다양한 노력이 이루어지고 있으며 에너지 사용량을 예측하는 연구도 활발하게 이뤄지고 있다.

우리나라에 편의점이 도입된 것은 1989년 5월 서울 올림픽선수촌에 세븐일레븐 1호점이 개점하면서 부터이다(백인수, 2006). 이후 다양한 브랜드의 편의점이 생겨났으며, 1인 가구의 증가와 팬더믹 확산으로 점포 수는 계속적으로 증가하고 있다. 산업통상자원부와 업계에 따르면, 2022년 2월 전국 편의점 수는 4만 2,672개로 2017년 3만 4,465개보다 8,207개 증가하였다(아시아경제, 2022).

편의점은 에너지 소비 측면에서 다른 유통 채널과 다르게 하루 24시간 비교적 일정하게 에너지를 사용하며 타 업종이 비해 냉동 및 냉장 에너지 비중이 크다(이동호 외, 2020). 만약 편의점에서 실제 소비되는 전력 데이터를 기반으로 최적의 전력 사용량을 예측할 수 있다면 보다 효율적인 에너지 소비 방안을 찾을 수 있을 것이다.

본 연구는 머신러닝을 이용해서 편의점의 전력 사용량을 예측할 수 있는 모델을 구축하고, 이를 이용해 계약전력을 최적화하는 방안을 제안하고자 한다. 이를 위해 전국의 특정 브랜드 편의점(이하 본문에서는 ‘A편의점’으로 칭한다)으로부터 수집한 실제 전력량 데이터를 이용하였다. XGBoost(eXtreme Gradient Boost) 모델을 이용하여 지점별 최적 계약전력 예측값을 찾아내고 이 과정에서 모델을 구성하는 중요 변수들에 대해서도 고찰하고자 한다.

2. 이론적 배경

2.1 XGBoost

XGBoost는 최소한의 리소스를 사용하여 대규모

머신러닝 모델로 확장 가능한 트리 부스팅 시스템이다. 2015년 Kaggle 블로그에 게시된 29개의 챌린지 우승 솔루션 중 17개 솔루션이 XGBoost를 사용, KDDCup 2015에서도 XGBoost는 상위 10위 안에 드는 모든 팀이 사용할 정도로 높은 성능을 나타내고 있다(Tianqi and Carlos, 2016). XGBoost는 데이터 처리 시간이 빠르고 과적합 규제(Regularization) 부재 문제를 해결하여 예측 성능이 높다는 장점이 있다(Shi et al., 2019). 실제로 Li and Zhang(2020)은 정형외과 분야에서 3가지 형태의 분류예측 모델을 구성하고 성능을 비교 분석한 결과, XGBoost 알고리즘 기반의 분류예측 모델이 더 높은 정확도와 더 빠른 계산 속도를 나타냄을 실증한 바 있다.

XGBoost의 그래디언트 트리 부스팅(Gradient Tree Boosting)은 그래디언트 부스팅 머신(GBM) 또는 그래디언트 부스팅 회귀 트리(GBRT)라고도 한다. 트리 앙상블 모델에서 Classification and Regression Tree(CART)를 학습하는데 정규화된 손실 함수(Regularized Loss Function) l 를 사용하고, 해당 목적함수 L 을 최소화하는 방법은 식 (1)과 같다.

$$\mathcal{L}(\mathcal{O}) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k), \quad (1)$$

$$\text{where } \Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (2)$$

l 은 실제 output y_i 와 예측 output \hat{y}_i 사이의 차이를 계산하는 미분 가능한 Convex 손실 함수이다. 식 (2)에서 각 Tree에 대하여 Ω 라는 정규화 매개변수를 두었다. 전통적 그래디언트 트리 부스팅에서는 이러한 정규화 매개변수가 없다. 반면, XGBoost에서는 보다 간단하고 오버피팅 방지 모델을 구성하기 위해 정규화 매개변수를 추가함으로써 오버피팅을 방지할 수 있다. XGBoost는 병렬 및 분산 컴퓨팅을 사용하면 학습 속도가 빨라져 모델 탐색이 더 빠르고, 확장성이 뛰어난 종단 간 트리 부스팅 시스템 설계와 구축이 용이한 장점이 있다. 또한 회귀모델과 분류모델 모두 활용할 수 있다(Tianqi and Carlos, 2016).

실제로 XGBoost 모델의 회귀 또는 분류모델에 대해 다양한 분야에서 연구가 수행되었는데, 황혜진 외(2018)는 XGBoost 분류 모델 해석을 통해 노인의 인지능력 개선, 악화 요인 및 중요 인자에 대해 탐색했고, 이용준, 선종환(2020)은 XGBoost를 활용 고속도로 콘크리트 포장 열화 예측을 위해 분류 학습을 기반으로 하는 파손 예측 방법을 제안했으며, 최성현, 허진(2020)은 하이퍼파라미터를 최적화하여 XGBoost 회귀 모델을 활용한 태양광 출력을 예측한 바 있다. 또한 오재영 외(2019)는 전력거래소 입장에서 전국 규모의 단일 수요량을 예측했는데, 이때 XGBoost 시계열 회귀 기법을 활용한 단기 전력 수요 예측과 하이퍼파라미터 변화에 따른 영향 분석을 수행하였다.

2.2 선행연구

머신러닝과 딥러닝 연구 이전에는 지능형 전력망을 통해 전력량을 예측하는 연구가 활발하게 진행되었다. 관련하여 이은지 외(2012)는 IT 최적화 기술을 이용해 전력량을 파악하여 전력 수요를 제어하는 기법을 제안하였다. 그러나 최근 들어서는 머신러닝과 딥러닝을 기법을 활용하여 에너지 사용량을 정밀하게 예측하는 방안을 찾는 연구가 다양한 분야에서 진행되고 있다. 이주희와 이강운(2021)은 일반 가구를 대상으로 기기별 에너지 사용량 예측에서 스마트 플러그로부터 6개월간의 시계열 측정 데이터를 DB 화시켜 R^2 로 LSTM(Long Short Term Memory) 모델의 성능을 확인했으며, 전병기 외(2019)는 1인 가구 한 세대를 대상으로 전기 수요 예측에 대해 40일 동안의 시간별 전기에너지 사용량 실측 및 재실 정보를 반영하여 MBE, RMSE, CVMSE로 LSTM 모델의 성능을 확인하였다. 오하령 외(2022)는 유통 분야를 대상으로 국내 대형 할인점 1곳의 2년 6개월간 스마트 전력량계 데이터를 이용하여, 다중 모델 머신러닝(GRU, LSTM, RNN 모델)을 구축하여 MAPE로 예측 정확도를 확인했으며, 성종훈, 조영식(2019)은 공장에서 사용되는 에너지 사

용량 데이터 분석에 대해, 생산설비 데이터를 이용 군집분석을 통해 에너지 소비 패턴과 공장 내의 장비 및 시설의 사용 패턴을 도출하였다. 문지훈 외(2016)는 서울 소재 대학의 강의실과 연구실로 구성된 건물을 대상으로 1년 6개월의 데이터에 ANN과 SVR을 적용하여 MAE, MAPE, RMSE로 예측 전력 사용량의 성능을 확인하였다.

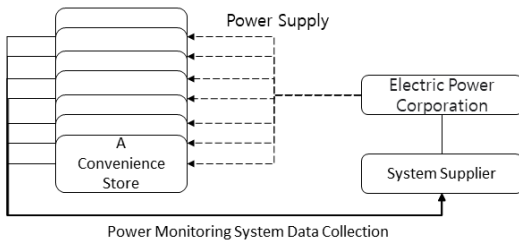
본 연구와 유사한 편의점 전력 사용량에 관한 연구로는 이동호 외(2020)가 특정 편의점 1곳에 대해 289일간 내부 설비들의 전기사용량을 직접 제작한 모듈로부터 수집한 데이터를 이용하여 날씨와의 상관관계에 관해 연구한 바 있다. 연구자는 1개 편의점만을 측정하여 냉장 및 냉난방 전력 사용량이 기온과 상관관계가 있다는 것을 규명하였다. 해외 유사 연구로 대만의 1,052개 편의점에 대한 설문 조사 데이터를 이용해 편의점 에너지 소비 성과를 측정하는 연구가 있었으나(Kuo et al., 2018) 해당 연구는 설문조사 데이터가 사용되었다는 점에서 계측기와 시스템을 통해 수집된 데이터를 이용하는 금번 연구와는 차이가 있다. 그 외 전력 사용 데이터를 활용하여 전기기기 소음의 불편함을 평가한 연구도 있다(Ryu et al., 2020). 이처럼 대형마트나 공장, 교육기관, 편의점의 전력 데이터를 분석하여 예측 성능을 파악하고 사용 패턴을 도출한 연구는 일부 확인되나, 전국 단위의 데이터로 전력량을 예측한 연구는 아직까지 확인되지 않고 있다.

3. 연구 방법

3.1 연구문제

현재 전국의 A편의점은 전력회사와 편의점주 간에 편의점 사용 전력량에 대해 연 단위 또는 6개월 단위로 계약을 체결한다. 계약기간 동안 계약전력을 초과해서 사용하면 편의점이 페널티를 물어야 하지만, 계약전력 내에서 사용하는 경우에는 계약전력 대비 적은 사용을 하더라도 편의점이 따로 보상받는 시스템은 아니다. [그림 1]과 같이 전국의 A편의점

에 전력 모니터링 시스템을 공급하고, 유지보수 및 전력계약 갱신업무를 수행하는 기업이, 전력 모니터링 시스템으로부터 수집되는 전력 소비량을 포함한 편의점 전력 관련 데이터를 보유하고 있다. 전력 모니터링 시스템의 공급기업은 축적된 데이터를 기반으로 매년 또는 6개월마다 전력 사용량에 대한 계약을 갱신하는 과정에서 해당 편의점에 최적화된 전력 사용 예측값을 제공하고자 한다.



[그림 1] 편의점 전력 모니터링 시스템 운영 환경

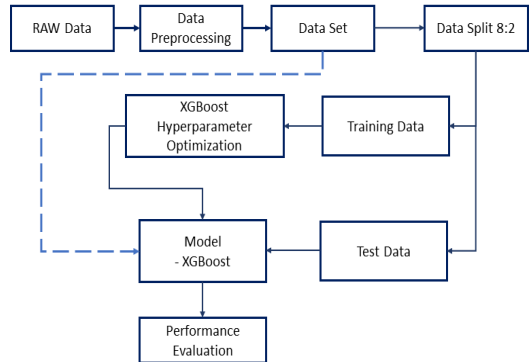
전력 사용량 예측 모델을 통해 전력 사용량이 많이 필요한 점포에는 최적 전력 사용량과 페널티 비용 절감 방안을 제안할 수 있고, 전력 사용량의 적은 편의점에도 좀 더 낮게 최적화된 계약전력을 제안할 수 있을 것이다. 더불어, 개점을 준비하는 점포에 대한 최적화된 최초 계약전력을 제안하는 방안도 찾고자 한다. 이를 위해 본 연구는 아래와 같이 연구문제를 정의한다.

- 연구 문제 1: 기존 매장에 대한 갱신용 계약 전력량의 예측
- 연구 문제 2: 신규 매장에 대한 제안 가능한 최초 계약 전력량의 예측

3.2 자료의 구성

본 논문에서의 분석 방법에 대한 분석 프레임워크는 [그림 2]와 같이 구성하였다. 제공된 원시 데이터의 전처리 과정을 통해, 결측치 및 이상치를 제거하고, 분석용 데이터 셋을 구축한다. 여기서, 결측치 및 이상치 제거에 따른 영향도를 확인하기 위

해서 결측치 및 이상치 모두 0으로 대체하는 방법, 결측치 및 이상치 모두 평균값으로 대체하는 방법, 결측치는 0으로 대체하고 이상치는 그대로 두는 방법의 3가지 Case로 구분하여 진행하고자 한다. 구축된 3가지 분석용 데이터 셋을 각각 무작위 추출 방법으로 학습 데이터 셋과 시험 데이터 셋으로 분할한다. 분할된 학습 데이터를 이용해 XGBoost 하이퍼파라미터 튜닝을 통해 최적화 파라미터를 구축하고, 시험 데이터를 이용해 구축된 XGBoost 모델에 대한 성능 평가를 수행한다. 그 다음 3가지 Case에 따른 분석용 데이터 셋 전체를 앞서 최적화한 각각의 XGBoost 모델에 적용함으로써 각 Case 별 성능을 예측하고, 시각화 과정을 통해 각각의 XGBoost 모델 성능에 영향을 미치는 중요 변수들을 확인해 본다.



[그림 2] 제안된 분석 절차

본 논문에서는 XGBoost 회귀 모델의 예측성능을 평가하기 위해 평균제곱오차(MSE, Mean Square Error), 평균절대오차(MAE, Mean Absolute Error), 평균제곱근오차(RMSE, Root Mean Squared Error), 결정계수인 R^2 및 상관도를 각각 평가지표로 사용하였다.

평균제곱오차란 잔차의 제곱에 평균을 취한 값으로 개별 관측값들이 중심에서 얼마나 멀리 떨어져 있는지의 상대 척도를 나타내는데, 값이 작을수록 추정의 정확성이 높다. 평균제곱오차는 아래 수식 (3)과 같이 표현된다.

$$\text{MSE} = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n} \quad (3)$$

평균절대오차란 잔차에 절대값을 취하고 평균을 낸 것으로 직관적으로 확인할 수 있다는 장점이 있으며, 아래 식 (4)와 같이 표현된다.

$$\text{MAE} = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{n} \quad (4)$$

평균제곱근오차는 잔차의 제곱에 대해 평균을 취한 값에 루트를 씌운 값으로 오류 지표를 실제값과 유사한 단위로 다시 변환하여 쉬운 해석을 지원한다. 평균제곱근오차는 아래 식 (5)와 같이 표현된다.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (5)$$

회귀분석에서 독립변수가 종속변수를 얼마나 설명해주는지를 나타내는 결정계수인 R^2 값도 성능검증 지표로써 활용하였다. R^2 의 수식은 아래 식 (6)과 같이 표현된다. 일반적으로 회귀모형에서 독립변수가 많은 회귀식에서는 수정된 결정계수(Adjusted R^2)를 사용하기도 한다. 그러나 본 연구에서는 수정된 결정계수를 사용하지 않고 상대 비교를 위해 R^2 만 사용하기로 한다.

$$R^2 = \frac{\text{SSE}}{\text{SST}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

위에서 기술한 MSE, MAE, RMSE는 모델마다 결과 값이 다르고, 0에 근접한 값으로 표현되면 좋은 성능을 발휘하지만, 0에 얼마나 근접을 해야 우수한 성능으로 판단되는지에 대해 성능 판단에 어려움이 있을 수 있다. 이에 본 연구에서는 계약전력과 예측된 계약전력과의 피어슨 상관계수를 통해 예측된 계약전력이 계약전력과 얼마나 상관이 있는지에 평가하고자 한다. 상관도는 아래 식 (7)과 같이 표현된다.

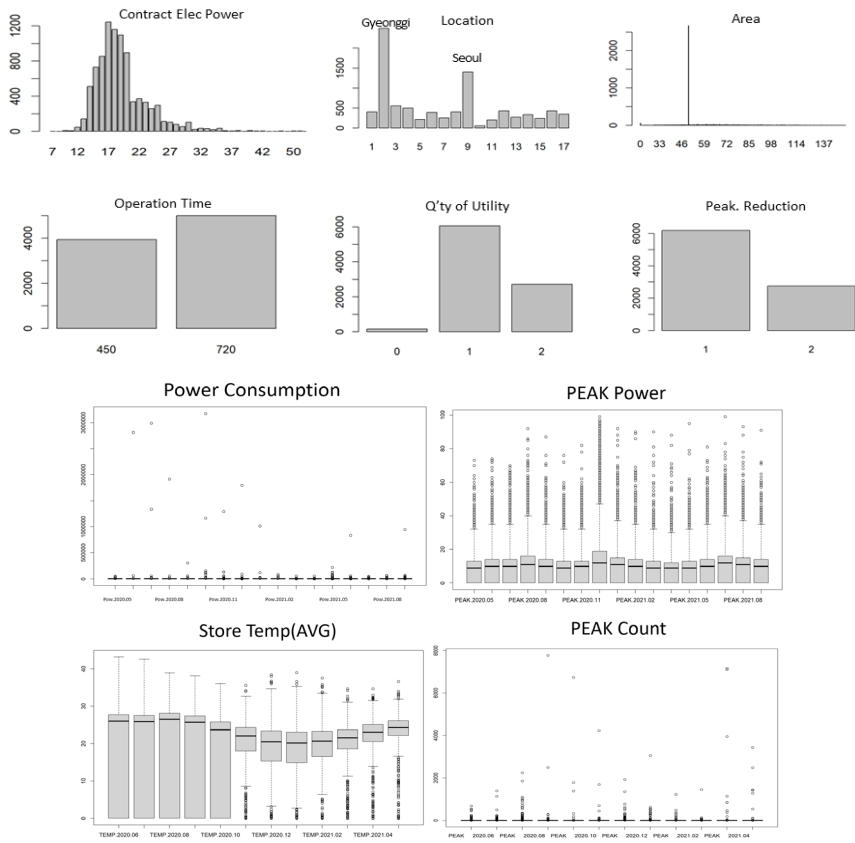
$$\text{Corr.} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}} \quad (7)$$

3.3 데이터 전처리

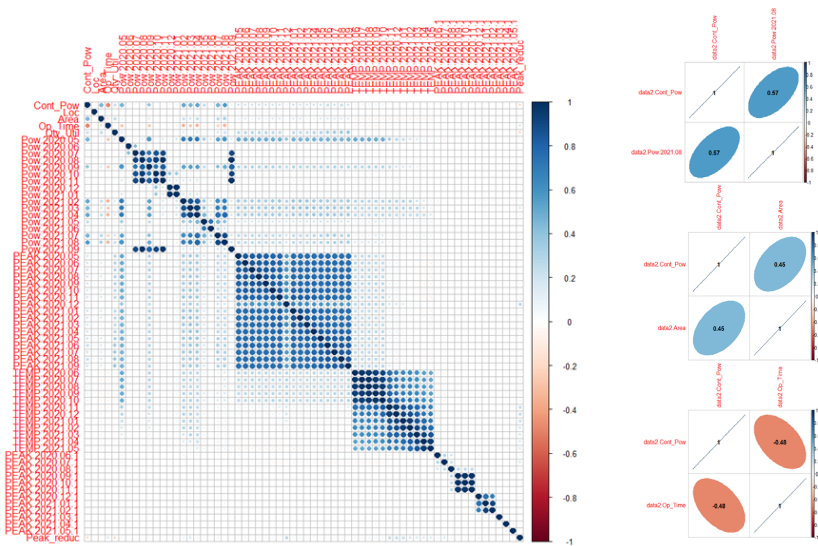
본 연구의 원천 데이터는 전국의 A편의점에 구축된 전력 모니터링 시스템으로부터 수집된 전국 편의점 전력 사용량으로, 2020년 5월부터 2021년 9월까지의 점포명, 계약전력, 설치지역, 면적, 운용시간, 냉난방기대수, 전력사용량, 최대피크, 매장실내온도, 피크제어횟수, 피크제어1회당 평균감소량, 실외온도 등으로 구분된 1년 4개월(16개월) 분의 데이터이다. 원천 데이터는 전국의 A편의점에 구축된 전력 모니터링 시스템으로부터 수집된 전국 편의점 전력 사용량으로, 2020년 5월부터 2021년 9월까지 9,701개의 점포 데이터와 68개의 변수로 구성된 csv 파일을 제공받아, 결측치 등을 제거한 1차 정제 데이터는 8,935개의 점포 데이터와 68개의 변수로 구성된 원시(RAW) 데이터로 분석을 수행했다. [그림 3]의 1차 정제 데이터는 8935행 × 68개의 변수로 구성되며, 행에 해당하는 전국 편의점 수는 8935개로 중복 데이터는 없는 것으로 확인하였다. 본 논문에서는 이렇게 확보한 1차 정제 데이터를 원시 데이터로 삼아 이하 연구를 진행한다.

원시 데이터를 기준으로 코드화(형 변환)를 비롯한 추가 전처리를 진행하였다. 설치지역 컬럼에 대해서는 <표 1>과 같이 문자형 데이터에 숫자형 코드를 부여하고, 피크감소량 컬럼도 공란은 1로, 400W~600W라는 값들은 2로 코드를 부여하였다. 실외온도 데이터는 측정치가 모두 결측치라서 해당 컬럼을 삭제하였다.

추가적인 데이터 전처리로 NA 결측치와 이상치 제거 방법을 <표 2>와 같이 3가지로 구분하여 진행하였다. 다만 XGBoost 모델이 식 (1)과 같이 정규화 항을 포함하고 있어서, 데이터 셋 전처리에 대해 별도의 정규화 과정은 거치지 않았다. 이상의 전처리 과정을 거쳐 아래와 같이 3가지 Case의 분석용 데이터 셋을 구축하였다. Case 1은 결측치와 이상



[그림 4] RAW 데이터의 빈도분석 결과



[그림 5] 결측값 제거 후 상관관계 분석(히트 맵)

나머지 Case에 대해서도 빈도분석 및 상관분석을 수행한 결과, 각 Case 별로 특정 월의 전력 사용량 데이터에서는 차이가 나지만, 면적 및 운영시간에는 차이가 없는 것으로 나타났다. 다만, 계약전력과 전력 사용량에 대한 상관계수에 근소한 차이를 확인하였는데, 그 결과는 <표 3>과 같다.

<표 3> 이상치 제거 후 계약전력-소비전력 상관관계 분석

CASE	전력 소비	상관관계
Case 1	Pow.2020.09	0.52
Case 2	Pow.2021.09	0.59
Case 3	Pow.2021.08	0.57

4.3 하이퍼파라미터 최적화

XGBoost의 하이퍼파라미터 조정은 XGBoost 모델의 전체 기능을 구동하는 일반 파라미터, 부스터의 성능을 제어하는 부스터 파라미터, 부스터의 학습 프로세스의 설정 및 평가에 대한 학습작업 파

라미터 등 3가지 범주로 구분할 수 있다. 본 연구에서는 부스터 파라미터를 최적화할 목적으로 앞서 구분한 3가지 Case 별로 하이퍼파라미터를 최적화하였으며, 여기에 GridsearchCV를 이용하였다. 튜닝 대상 하이퍼파라미터의 유형과 각 튜닝값은 사전 모델 분석에서 이와 같은 범위의 파라미터를 사용했을 경우 상관도가 94% 이상 나온 값들로 <표 4>와 같이 설정하였다. 일부 하이퍼파라미터에 대해서는 최적화 시간을 단축하기 위해 디폴트 값으로 설정하였다. 한편, GridsearchCV의 교차검증 폴드 수는 보다 더 정밀한 검증을 위해 7로 설정하고 최적화를 진행하였다.

원시 데이터로부터 전처리 된 3가지 분석용 데이터셋은 XGBoost 분석을 위해 8:2의 학습 데이터와 시험 데이터로 분할하여 Matrix 형태로 변환하였다. GridsearchCV에서 학습 데이터를 XGBoost 모델로 구축하여 격자(grid) 구조로 <표 5>와 같이 81개 파라미터 조합으로 학습 및 교차검증을 수행하였다. 교차검증을 마친 최종 최적화된 파라미

<표 4> 하이퍼파라미터 유형, 정의, 조정값

하이퍼파라미터	정의	조정값
max_depth	The maximum depth of a tree.	9, 10, 11
nrounds	It controls the maximum number of iterations.	100, 150, 200
eta	It is the step size shrinkage used in update to prevent overfitting.	0.3, 0.4, 0.5
colsample_bytree	The subsample ratio of columns when constructing each tree.	0.8, 0.9, 1.0
gamma	Specifies the minimum loss reduction required for splitting.	0 = Default
subsample	It denotes the fraction of observations to be randomly samples for each tree.	1 = Default
min_child_weight	The minimum sum of weights of all observations required in a child.	1 = Default

<표 5> 하이퍼파라미터의 그리드 유형

No.	max_depth	nrounds	eta	gamma	subsample	min_child_weight
1	9	100	0.3	0	1	1
2	10	100	0.3	0	1	1
3	11	100	0.3	0	1	1
⋮						
79	9	200	0.5	0	1	1
80	10	200	0.5	0	1	1
81	11	200	0.5	0	1	1

터가 설정되면 해당 파라미터를 반영한 모델에 시험 데이터를 적용하여 성능평가를 수행하였다. 그리고, GridsearchCV는 시험 데이터로 MSE 값이 가장 낮은 하이퍼파라미터 조합을 최적의 하이퍼파라미터로 선정하였다.

4.4 분석 결과

분석 결과, 최적 하이퍼파라미터는 <표 6>과 같이 3가지 Case에 따라 차이를 보였다. XGBoost 모델의 트리 구성을 위한 최대 깊이인 max_depth는 Case 별로 모두 다르게 나타났다. nrounds 수는 Case 1이 150회, 나머지 Case 2와 Case 3은 100회로 나타났다. 그 외 과적합 방지를 위한 스텝단계 축소값인 eta가 Case 2가 0.2, 나머지는 0.3으로 결정되었다. colsample_bytree 값은 Case 2, 3은 동일한 0.9이며, Case 1은 0.8로 나타났다. 그 외 하이퍼파라미터는 Default로 설정된 값이다. 결론적으로 Case 1, 2, 3은 모두 서로 다른 최적값의 하이퍼파라

미터를 갖게 되었으며, 이에 따라 Case 별 XGBoost 회귀 모델을 구성하였다.

각 Case별 XGBoost 회귀 모델 분석용 데이터 셋 전체를 적용하여 <표 7>과 같은 최종 성능평가 결과가 도출되었다. 전반적으로 Case 1이 Case 2 또는 Case 3에 비해 근소하게 높은 성능을 보이는 것으로 나타났다.

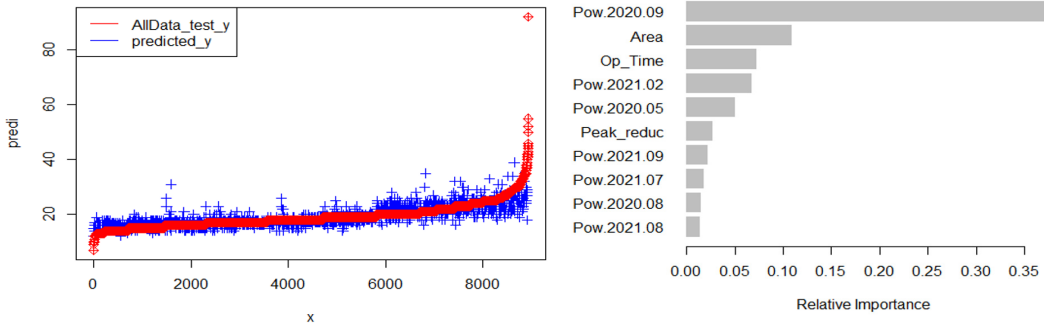
회귀식의 설명력을 나타내는 R^2 값이 가장 높은 Case 1은 0.931로 나타났으며, 가장 낮은 Case 2는 0.919이다. 상관도(Corr.)의 경우, 가장 높은 Case 1은 0.965이고, 가장 낮은 Case 2가 0.959로 나타나 수치상으로 보면 Case 1의 상관도가 Case 2에 비해 0.59 % 더 높은 것으로 나타났다. 그리고 각 Case 별 XGBoost 모델에서 가장 중요하게 작용하는 변수로서 Case 1은 2020년 9월 전력 사용량, 면적, 운영시간으로 나타나고, Case 2는 2021년 2월 전력 사용량, 2020년 9월 전력 사용량, 운영시간으로 확인이 되며, Case 3은 2020년 9월 전력 사용량, 2021년 2월 전력 사용량, 운영시간 순으로 각각 확인되

<표 6> 하이퍼파라미터 최적화 결과

HyperParameter	Case 1	Case 2	Case 3
max_depth	10	9	11
nrounds	150	100	100
eta	0.3	0.2	0.3
colsample_bytree	0.8	0.9	0.9
gamma	0	0	0
subsample	1	1	1

<표 7> 케이스 별 성능평가 결과

Evaluation Index	Case 1	Case 2	Case 3
MSE	1.293311	1.505592	1.335542
MAE	0.3551575	0.5771986	0.3688141
RMSE	1.137239	1.227026	1.155656
R^2	0.9310056	0.919681	0.9287527
Corr.	0.9649479	0.9590911	0.9637697
Important Variable 1	Pow.2020.09	Pow.2021.02	Pow.2020.09
Important Variable 2	Area	Pow.2020.09	Pow.2021.02
Important Variable 3	Op Time	Op Time	Op Time



[그림 6] Case 1의 최종 예측 결과 및 핵심 변수 시각화

었다. [그림 6]은 Case 1에 대한 계약전력과 예측된 계약전력과의 관계에 대한 시각화와 Case 1의 XGBoost 모델에서 중요 변수 10개에 대한 결과를 시각화한 내용이다.

결과적으로 본 논문에서 제기한 연구문제 1(기존 매장에서의 향후 재계약 전력을 위한 예측)에 대해서는 연구진이 구축한 XGBoost 모델로 예측이 가능하다는 결론을 도출하였다. 하지만 연구문제 2(신규 매장에 대한 제안 가능한 계약전력량 예측)에 대해서는 예측이 어려울 것으로 판단된다. 그 이유는 <표 7>에서 각 Case 별 XGBoost 모델에서 가장 중요한 변수로 도출된, 전력 사용량과 운영시간은 실제 편의점을 운영했을 때만 수집 가능한 데이터로 개점 이력이 없는 매장에서는 확보할 수 없기 때문이다.

5. 결 론

본 연구는 전국 A편의점에서 실시간 수집되는 데이터를 이용하여 1) 계약 갱신을 위한 계약전력량 예측값과 2) 개점 전(신규) 편의점에 적합한 계약전력량 예측값을 제안할 목적으로 전력량 예측 모델을 구축하였다. 이를 위해 XGBoost 회귀 모델을 작성하였으며, 최적의 하이퍼파라미터도 도출하였다. 모델의 구축과정에서 결측치 및 이상치의 처리 방법을 3가지 Case로 구분해 예측 성능에 어떤 영향을 미치는지도 확인하였다.

구축된 예측 모델을 통해 갱신을 위한 계약전력 예측(연구문제 1)이 가능함을 확인하였다. 그리고 각 Case 별 결측치 및 이상치 처리는 예측 결과에 큰 영향을 미치지 않는 것으로 파악되었다. 이는 결측치 및 이상치 처리 방법에 따라 데이터 셋이 달라지더라도 하이퍼파라미터 최적화 과정으로 각각에 최적화된 모델이 구축된다는 것을 의미한다. 또한 구축한 예측 모델에서의 중요 변수로 특정 사용 월의 전력 사용량과 편의점의 면적, 운영시간 등이 가장 영향을 많이 주는 변수라는 것도 확인할 수 있었다. 하지만 개점 전의 신규 편의점에 대한 계약전력 제안(연구문제 2)은 어렵다는 결론을 얻었다.

금번 연구는 점포별 전력수요 실적을 토대로 각 지점의 전력 수요를 예측한다는 점에서 기존 유사 연구들과 차별성을 가지고 있다. 즉, 기존 연구들이 단일 지점의 과거 시계열 데이터를 토대로 전력수요를 예측하거나, 전국 데이터를 사용하더라도 개별 지점이 아닌 전국 규모의 수요를 예측하는 데 머무르는 것과 달리 본 연구는 전력수요를 결정하는 변수들을 파악함으로써 지점별 수요를 예측하는 데 머신러닝을 적용했다는 점에서 연구 의의를 찾을 수 있겠다.

본 연구는 착수과정에서 XGBoost 외에 몇 가지 기법을 대상으로 기본적인 학습 소요시간을 간략히 비교한 자료만으로 XGBoost 모델을 선택해서 분석을 수행했기 때문에 다른 기법들과의 보다 정밀한

비교에 따른 결과를 확인하기 어렵다는 한계가 있다. 또한 데이터 수집 기간이 5월부터 이듬해 9월까지로 여름을 두 번 거치면서 총 전력 사용량에 영향을 미쳤을 것으로 예상된다. 전력 사용량에 대해 평균, 분산, 표준편차를 사용한 파생변수를 만들지 않고 각 사용 월에 대한 전체 데이터를 학습시켰기 때문에 총 전력량에는 영향을 주지 않았을 것이나, 개별 월별 사용량 예측에는 영향을 미칠 수 있을 것이다. 연구진은 데이터가 많을수록 더 높은 정확도를 나타낼 것이라는 판단으로 기간을 년 단위로 제한하지 않았는데, 이는 본 연구의 한계로 지적될 수 있겠다. 또한 성능지표의 경우 사용자 입장에서 직관적으로 이해하기 쉬운 지표를 선정하였는데, 보다 다양한 성능지표에 대해서도 더욱 심도 있는 분석이 이뤄질 필요가 있을 것이다. 그리고 결측치 및 이상치를 보다 섬세하게 처리할 수 있는 방법론을 적용할 수 있다면 예측 모델의 정확도 또한 좀 더 높일 수 있을 것으로 기대한다.

참고문헌

- 문지훈, 전상훈, 박진웅, 최영환, 황인준, “인공 신경망과 지지 벡터 회귀분석을 이용한 대학 캠퍼스 건물의 전력 사용량 예측 기법”, *정보처리학회논문지*, 제5권, 제10호, 2016, 293-302.
- 백인수, “한국 편의점의 성장 프로세스와 성장 전략”, *경영사연구*, 제42호, 2006, 159-187.
- 성종훈, 조영식, “머신러닝 기법을 활용한 공장 에너지 사용량 데이터 분석”, *정보처리학회논문지*, 제8권, 제4호, 2019, 87-92.
- 아시아경제, 2022, July 31, <https://www.asiae.co.kr/article/2022042507301358635>.
- 오재영, 함도현, 이용건, 김기백, “XGBoost 기법을 이용한 단기 전력 수요 예측 및 하이퍼파라미터 변화에 따른 영향 분석”, *전기학회논문지*, 제68권, 제9호, 2019, 1073-1078.
- 오하령, 임정현, 성영락, “다중 모델 머신러닝 기반 전력사용량 예측 정확도 향상 기법”, *전기학회논문지*, 제71권, 제6호, 2022, 876-883.
- 이동호, 김성환, 표성민, “편의점 냉동/냉장/냉난방 전력사용량 측정 및 분석”, *전기학회논문지*, 제69권, 제4호, 2020, 298-303.
- 이용준, 선종완, “XGBoost를 활용한 고속도로 콘크리트 포장 파손 예측”, *한국건설관리학회 논문집*, 제21권, 제6호, 2020, 46-55.
- 이은지, 서유리, 윤소영, 장혜린, 반효경, “IT 최적화 기술을 이용한 지능형전력망 환경의 스마트 빌딩 전력 스케줄링”, *한국IT서비스학회지*, 제11권, 제1호, 2012, 41-50.
- 이주희, 이강윤, “가구당 기기별 에너지 사용량 예측을 위한 딥러닝 모델의 설계 및 구현”, *한국빅데이터학회학회지*, 제6권, 제1호, 2021, 127-132.
- 전병기, 김의중, 이경호, 공민석, 신영기, “딥러닝을 이용한 재실정보 기반 건물의 전기 수요 예측 모델”, *설비공학논문집*, 제31권, 제1호, 2019, 22-31.
- 황혜진, 김수현, 송규원, “XGBoost 모델 해석을 통한 노인의 인지능력 개선·악화 요인 탐지”, *한국차세대컴퓨팅학회논문지*, 제14권, 제3호, 2018, 16-24.
- 최성현, 허진, “최적화 하이퍼파라미터의 XGBoost 학습자 기반 배깅 모델을 활용한 태양광 출력 예측”, *전기학회논문지*, 제69권, 제7호, 2020, 978-984.
- Kuo, C. F., C. H. Lin, and M. H. Lee, “Analyze the energy consumption characteristics and affecting factors of Taiwan’s convenience stores—using the big data mining approach”, *Energy and Buildings*, Vol. 168, 2018, 120-136.
- Li, S. and X. Zhang, “Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm”, *Neural Computing and Applications*, Vol.32, 1971-1979.
- Ryu, D. H., R. H. Kim, S. H. Choi, K. J. Kim, Y. M. Ko, Y.-J. Kim, M. Song, and D. G.

- Choi, “Utilizing Electricity Consumption Data to Assess the Noise Discomfort Caused by Electrical Appliances between Neighbors: A Case Study of a Campus Apartment Building”, *Sustainability*, Vol.12, No.20, 2020, 1-16.
- Shi, X., Y. D. Wong, M. Z. F. Li, C. Palanisamy, and C. A. Chai, “Feature Learning Approach Based on XGBoost for Driving Assessment and Risk Prediction”, *Accident Analysis & Prevention*, Vol. 129, 2019, 170-179.
- Tianqi, C. and G. Carlos, “XGBoost: A Scalable Tree Boosting System”, *KDD '16(Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785-794.

◆ About the Authors ◆



김 상 민 (randonnee20@naver.com)

금오공과대학교 대학원 컨설팅학 박사학위를 취득하고, 현재 (주)임픽스 기술 연구소에서 제조 빅데이터 및 AI 컨설팅을 수행 중이다. 관심분야는 제조 빅데이터, ML, DL이다.



박 찬 권 (chankwon@hycu.ac.kr)

서울대학교 산업공학과에서 박사학위를 취득하고 현재 한양사이버대학교 생산물류유통학과에 재직하고 있다. 관심분야는 SCM, 의사결정, 최적화이다.



이 지 은 (scully1215@hycu.ac.kr)

한양대학교 정보통신대학원 정보기술경영학과를 졸업하고 현재 한양사이버대학교 경영정보·AI비즈니스학과에 재직하고 있다. 관심분야는 에듀테크, 지식경영, 인공지능이다.