



ISSN: 3022-5388

JKAI website: <https://accesson.kr/jkaia>DOI: <http://dx.doi.org/10.24225/jkaia.2023.1.2.21>

# 회귀 모델을 활용한 철강 기업의 에너지 소비 예측

## Forecasting Energy Consumption of Steel Industry Using Regression Model

Sung-Ho KANG, Hyun-Ki KIM<sup>2</sup>

Received: October 30, 2023. Revised: November 16, 2023. Accepted: December 30, 2023.

### Abstract

The purpose of this study was to compare the performance using multiple regression models to predict the energy consumption of steel industry. Specific independent variables were selected in consideration of correlation among various attributes such as CO2 concentration, NSM, Week Status, Day of week, and Load Type, and preprocessing was performed to solve the multicollinearity problem. In data preprocessing, we evaluated linear and nonlinear relationships between each attribute through correlation analysis. In particular, we decided to select variables with high correlation and include appropriate variables in the final model to prevent multicollinearity problems. Among the many regression models learned, Boosted Decision Tree Regression showed the best predictive performance. Ensemble learning in this model was able to effectively learn complex patterns while preventing overfitting by combining multiple decision trees. Consequently, these predictive models are expected to provide important information for improving energy efficiency and management decision-making at steel industry. In the future, we plan to improve the performance of the model by collecting more data and extending variables, and the application of the model considering interactions with external factors will also be considered.

**Keywords :** Machine Learning, Energy Consumption Prediction, Regression Models, Correlation Analysis

**Major Classification Code :** Artificial Intelligence, etc

## 1. Introduction

### 1.1. Significance of the Study

최근의 에너지 소비 예측은 산업 분야에서 핵심적인 경영 요소 중 하나로 인식되고 있다 (Kim & Hong, 2022). 특히 철강 산업과 같이 에너지 집약적인 분야에서는 정확한 에너지 소비 예측이 기업의 생산 계획, 에너지 효율성 향상, 환경 지속 가능성 등 다양한 측면에서 중요한 역할을 수행한다(Xiao, Shao, Liang, & Wang, 2016). 또한 최근 에너지 소비 패턴을 모델링하기 위해 여러 통계

및 인공 지능 기술이 개발되었다. 기계 학습 및 데이터 마이닝 기술을 사용하여 에너지 소비 예측을 조사하는 것은 산업 분야 전반에 걸쳐 확인할 수 있다. 데이터 마이닝 접근 방식은 현장에서 일반 사용자가 사용하기에 매우 유용하고 편리하며 많은 응용 분야에서 더 인기가 있으므로 본 연구에서 사용되었다 (Simeone, 2018).

본 연구에서는 Azure Machine Learning Studio 를 활용하여 철강 기업의 에너지 소비를 예측하는 과정을 다룬다. Azure Machine Learning Studio 는 머신러닝 모델의 개발, 학습, 평가, 배포를 간편하게 수행할 수 있는 강력한

1 First Author. Student, Big data medical conference, Eulji University, Korea. Email: ho021029@gmail.com

2 Corresponding Author or Second Author, 신남정보통신 대표. Email: r48019@naver.com

© Copyright: The Author(s)

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 noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

도구로, 이를 통해 철강 산업에서의 에너지 소비 패턴을 파악하고 더 효율적인 에너지 관리 방안을 모색하는 데 도움이 된다. 이 연구에서는 철강 기업 데이터셋의 소개부터 시작하여 어떤 변수들 중요하게 작용하는지를 선정하고, 그에 기반한 모델을 학습하고 평가하는 전체적인 프로세스를 살펴본다. 최종적으로, 에너지 소비 예측을 향상시키기 위한 실질적인 인사이트를 도출할 것이다. 이러한 분석은 기업의 에너지 효율성을 향상시키고, 생산 과정을 최적화하여 비용을 절감하며, 동시에 환경 지속 가능성을 고려하는 데 기여할 것으로 기대된다.

## 1.2. Related Research

관련 연구로 Sathishkumar et al.의 연구에서는 동일한 데이터셋을 활용하여 소규모 철강 산업의 에너지 소비를 예측하는 데 중점을 두었다. 주요 목적은 최적의 회귀 알고리즘을 통해 최소 오차로 에너지 소비를 예측하는 것이었다. 결과는 랜덤포레스트 (Random Forest) 모델이 다른 회귀 모델에 비해 예측의 상대평균제곱오차 (Root Mean Squared Error), 평균절대오차 (Mean Absolute Error), 평균절대비오차 (Mean Absolute Percentage Error), 및 변동계수 (Coefficient of Value)값을 향상시켰다는 것을 보여주었다. 그러나 이 연구는 단일 산업에 대한 분석으로 제한되어 있었고, 다양한 산업을 포함한다면 정보를 더 확장할 수 있을 것으로 기대된다. 또한, 산업의 에너지 소비와 관련하여 생산되는 제품, 건물의 기하학, 장비의 종류 등 추가 특성을 조사할 수 있다. 이 연구는 탐색적 분석과 예측 모델 개발 과정에서 흥미로운 결과를 제시하였다 (Sathishkumar, Shin, & Cho, 2021).

Cauwer et al.의 연구에서는 에너지 소비를 추정하기 위해 통계 모델과 물리적 개념을 결합하여 사용하는 방안을 제안했다. 해당 연구는 단일 산업에 대한 분석에서 나아가, 다양한 산업에 대한 적용 가능성을 열어두었다. 이를 통해 에너지 소비와 관련된 생산 특성, 건물의 기하학, 그리고 장비 종류 등에 대한 추가적인 특성을 조사할 수 있을 것이다(De Cauwer, Van Mierlo, & Coosemans, 2015).

Xuemei et al.는 퍼지 지원 벡터 머신 및 퍼지 C-평균 클러스터링을 사용하여 건물 냉각 부하를 예측하는 연구를 수행했다. 이 방법은 예측 수요의 조건을 설정하여 모델의 효과적인 성능을 향상시켰다. 이러한 방법은 다양한 기술

및 기후 조건에 대응하여 건물 에너지 소비를 더 정확하게 예측하는 데 기여했다(L. Xuemei, D. Yuyan, D. Lixing, and J. Liangzhong, 2010).

Zhao et al.는 철강 산업의 에너지 시스템에 대한 장기적인 예측을 위해 산업 주도 의미론 및 초기 데이터의 세분화를 통합하는 그랜슬러 컴퓨팅 접근을 제안했다. 중국의 한 철강 공장에서 실제 산업용 에너지 데이터를 사용하여 제안된 방법의 효율성을 평가하였으며, 다른 데이터 기반 방법보다 우수한 결과를 도출하였다(J. Zhao, Z. Han, W. Pedrycz, and W. Wang, 2015).

Dong et al.는 서포트 벡터 머신(Support Vector Machine)을 건물 에너지 소비 예측에 적용하여 소형 건물에서의 효율적인 에너지 소비 예측을 제시했다. 해당 연구는 다양한 기상 조건과 건물 특성을 고려하여 모델의 성능을 개선하였으며, 건물 에너지 효율을 향상시키는 데 기여했다(B. Dong, C. Cao, and S. E. Lee, 2005).

위와 같은 연구들처럼 에너지 수요 예측을 위한 다양한 연구가 끊임없이 수행되었다. 본 연구에서는 다수의 회귀 모델을 활용하여 에너지 소비를 예측하는데 그 목적을 두고 있다. 신경망 회귀 (Neural Network Regression), 선형 회귀 (Linear Regression), 푸아송 회귀 (Poisson Regression), 부스티드 의사결정 회귀나무 (Boosted Decision Regression), 의사결정 포레스트 회귀 (Decision Forest Regression), 베이저안 선형 회귀 (Bayesian Linear Regression) 등 6 가지 회귀 모델을 선택하여 학습시키고, 이들 중에서 어떤 모델이 최적의 성능을 보이는지를 평균 제곱 오차 (Mean Squared Error)를 통해 평가할 것이다.

## 2. Body

본 연구에서 사용된 데이터셋은 철강 기업에서 수집되었다. 이 기업은 다양한 종류의 코일, 강판, 및 철판을 제조하며, 해당 데이터셋은 한국전력공사 웹사이트(pccs.kepco.go.kr)에서 확인 가능하며, 일일, 월간 및 연간 에너지 소비 패턴에 대한 자세한 내용이 계산되어 표시되고 있다.

데이터셋의 변수는 다음과 같다. 날짜(Date)는 매월 1 일에 측정된 연속 시간 데이터를 나타낸다. 전력 사용량(Usage\_kWh)은 산업용 에너지 소비량을 표현하고, 후행 전력(Lagging)과 선행 전력(Leading)은 각각 현재의 지연된 반응력과 선행된 반응력을 나타낸다.

이산화탄소(CO2)는 연속적인 이산화탄소 농도(ppm)를 나타내며, 자정으로부터의 초 단위 시간(NSM)은 자정으로부터의 경과 시간(초)을 나타낸다. 주말/평일 상태(Week status)는 범주형으로 주말(0) 또는 평일(1)을 나타내며, 요일(Day of week)은 범주형으로 일요일부터 토요일까지 나타낸다. 부하 유형(Load Type)은 범주형으로 경량 부하, 중량 부하, 최대 부하를 나타낸다. 이 중 특정 독립 변수들은 예측하고자 하는 전력 사용량이 아닌 다른 독립변수와 상관관계가 높으면 다중공선성 문제를 불러일으킬 수 있다(Lee, 2003). 이를 해결하기 위해 상관계수를 검토하여 특정 독립 변수들을 선택적으로 제거하기로 하였다. 이러한 접근 방식은 모델의 일반화 성능을 향상시키며, 불필요한 변수들을 제거하여 모델의 복잡성을 줄이는 데 도움이 된다. [표 1]은 각 변수와 종속변수 간의 상관계수를 나타내며, 연속형 변수들은 피어슨 상관계수 (Pearson Correlation Coefficient), 그리고 다른 변수들은 스피어만 상관분석 (Spearman Correlation Coefficient)을 통해 평가되었다.

**Table 1:** Correlation coefficients between each variable and its dependent variable

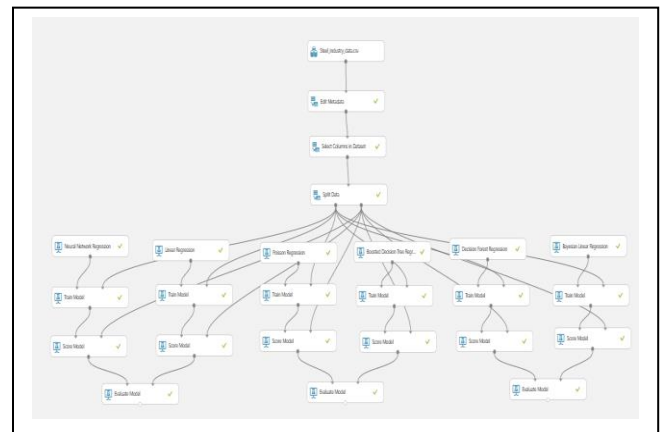
Pearson Correlation		Spearman Correlation	
Attribute	value	Attribute	Value
Lagging Current Reactive Power kVarh	0.9	NSM	0.32
Leading Current Reactive Power kVarh	-0.32	Week Status	-0.31
CO2(tCO2)	0.99	Day of week	-0.25
Lagging Current Power Factor	0.39	Load Type	0.7
Leading Current Power Factor	0.35	Season	0.13

상관계수 표를 분석한 결과, 피어슨 상관계수와 스피어만 상관계수를 통해 각 변수들 간의 선형 및 비선형 상관 관계를 확인할 수 있다. Lagging Current Reactive Power kVarh 변수와 CO2(tCO2) 변수가 에너지 소비량과 높은 상관관계를 보이는 것으로 나타났다. 그러나 두 변수 간의 상관계수는 0.9 로 매우 높게 나타났다. 이는 두 변수 간에 강한 선형 관계가 있을 가능성이 있으며, 이로 인해 다중공선성 문제가 발생할 수 있다. 다중공선성은 예측 모델에서 한 독립변수가 다른 독립변수와 강한 상관관계를 가질 때 발생하며, 모델의 불안정성과 해석의 어려움을 초래할 수 있다. 따라서, 다중공선성 문제를 피하기 위해

상관계수가 높은 Lagging Current Reactive Power kVarh 변수를 최종 모델에 포함시키지 않기로 결정하였다.

종속변수와 NSM 변수와는 0.32의 상관계수를 보였으며, Week Status 변수와는 -0.31, Day of week 변수와는 -0.25 로 각각 양의 상관관계 또는 음의 상관관계가 나타났다.

에너지 소비 예측 모델에 중요한 영향을 미치는 요소를 선택하기 위해 최종적으로 CO2 변수, NSM 변수, Week Status 변수, Day of week 변수, Load Type 변수를 예측 모델의 속성으로 선정하였다. 이러한 특성들은 날씨, 환경 변화, 주중/주말 여부, 요일, 부하 유형 등의 다양한 측면을 반영하여 에너지 소비를 예측하기 최적화되었다고 판단했기 때문이다. [그림 1]은 Azure Machine Learning Studio 에서의 머신러닝 과정을 시각화 했다.



**Figure 1:** Machine Learning Process

Azure Machine Learning Studio 를 활용하여 속성들을 선정하고 NSM, Week Status, Day of week, Load Type, Season 속성들은 연속형 변수가 아닌 범주형 변수로 간주되도록 메타데이터를 변환하였다. 그리고 학습 데이터와 검증 데이터를 8:2 비율로 나누어 신경망 회귀, 선형 회귀, 푸아송 회귀, 부스티드 의사결정 회귀나무, 의사결정 포레스트 회귀, 베이지안 선형 회귀를 사용하여 학습하였다.

[그림 2]에서 신경망 회귀 모델과 선형 회귀로 학습된 예측 모델을 검증데이터로 비교하여 시각화 했다.

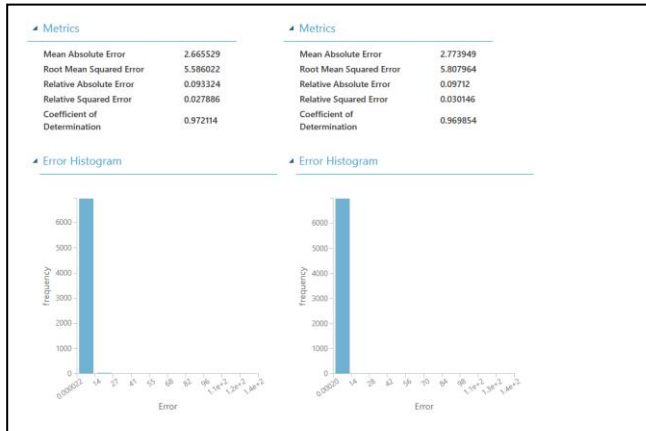


Figure 2: Neural Network Regression & Linear Regression

Figure 2 에는 신경망 회귀모델 및 선형 회귀 모델에 대한 예측 결과의 성능 지표가 제시되어 있다. 먼저, 신경망 회귀의 경우, 평균절대오차가 2.665529, 상대평균제곱오차가 5.586022 으로 나타났다. 상대절대오차 (Relative Absolute Error)와 상대제곱오차 (Relative Squared Error)는 각각 0.093324 및 0.027886 이며, 결정계수 (Coefficient of Determination)는 0.972114 로 나타났다.

이어서 선형 회귀 모델에서는 평균절대오차 가 2.773949, 상대평균제곱오차가 5.807964 로 측정되었다. 상대절대오차와 상대제곱오차는 각각 0.09712 및 0.030146 이며, 결정계수는 0.969854 로 나타났다. 두 모델 간의 성능을 비교할 때, 신경망 회귀 모델이 선형 회귀 모델보다 상대적으로 높은 결정계수 및 낮은 에러 지표를 보여 우수한 예측 능력을 나타내고 있다. [그림 3]에선 푸아송 회귀와 부스티드 의사결정 회귀나무로 학습된 예측 모델을 검증데이터로 비교하여 시각화 했다.



Figure 3: Poisson Regression & Boosted Decision Tree Regression

Figure 3 에서는 푸아송 회귀 및 부스티드 의사결정 회귀나무 모델에 대한 성능 지표가 제시되어 있다. 먼저, 푸아송 회귀 모델의 평균절대오차는 9.071456, 상대평균제곱오차는 16.356654 로 나타났다. 상대절대오차와 상대제곱오차는 각각 0.317604 및 0.239097 이며, 결정계수는 0.760903 로 나타났다. 부스티드 의사결정 회귀나무 모델에서는 평균절대오차가 2.408174, 상대평균제곱오차가 4.390733 으로 측정되었다. 상대절대오차와 상대제곱오차는 각각 0.084313 및 0.017229 이며, 결정계수는 0.982771 이다. 이러한 결과를 통해 부스티드 의사결정 회귀나무 모델이 푸아송 회귀 모델에 비해 낮은 에러 및 높은 결정계수를 보여주어 우수한 예측 능력을 갖고 있음을 확인할 수 있다.

[그림 4]에서는 의사결정 포레스트 회귀 & 베이지안 선형 회귀로 학습된 예측 모델을 검증데이터로 비교하여 시각화 했다.

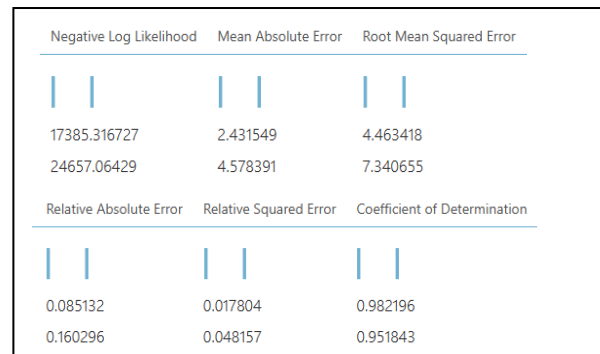


Figure 4: Decision Forest Regression & Bayesian Linear Regression

Figure 4 에서는 의사결정 포레스트 회귀 및 베이지안 선형 회귀 두 모델에 대한 다양한 성능 지표가 제시되어 있다.

의사결정 포레스트 회귀 모델의 경우, 음의 로그우도는 17385.316727, 평균절대오차는 2.431549, 상대평균제곱오차는 4.463418 로 나타났다. 상대절대오차와 상대제곱오차는 각각 0.085132 및 0.017804 이며, 결정계수는 0.982196 로 나타났다.

한편, 베이지안 선형 회귀 모델에서는 음의 로그우도가 24657.06429, 평균절대오차가 4.578391, 상대평균제곱오차가 7.340655 로 나타났다. 상대절대오차와 상대제곱오차는 각각 0.160296, 0.048157 이며, 결정계수는 0.951843 로 나왔다.

이를 통해 의사결정 포레스트 회귀가 베이지안 선형 회귀에 비해 상대적으로 더 낮은 에러 지표와 높은 결정계수를 보여 더 효과적으로 성능을 발휘하고 있음을 확인할 수 있다.

예측 모델을 비교한 결과 6 가지 회귀 모델 중 평균제곱오차(Mean Squared Error)가 가장 낮은 모델은 부스티드 의사결정 회귀나무 모델이었다.

### 3. Results

Azure Machine Learning Studio 를 활용하여 철강 기업의 에너지 소비를 예측하는 과정을 자세히 살펴보았다. 우선, 우리는 다수의 속성 중에서 CO2 농도, NSM, Week Status, Day of week, Load Type 등을 선정하여 이들을 독립변수로 설정했다. 특히, 상관계수 분석을 통해 각 속성 간의 상관관계를 평가하고, 이를 기반으로 선별한 독립변수들은 에너지 소비에 미치는 영향을 잘 반영하도록 고심했다.

데이터 전처리 과정에서는 다중공선성 문제를 방지하기 위해 일부 변수를 제거하였다. 또한, 예측 성능을 향상시키기 위해 특정 독립 변수들과 종속 변수 간의 상관계수를 고려하여 최종 독립 변수를 결정하였다. 이를 통해 모델의 복잡성을 줄이고, 불필요한 변수를 제거하여 예측 정확도를 향상시켰다.

여러 회귀 모델을 학습시켜본 결과, 부스티드 의사결정 회귀나무 모델이 다른 모델에 비해 우수한 성능을 보였다. 이는 앙상블 학습의 특성으로 여러 결정 트리를 조합해 과적합을 방지하면서도 복잡한 패턴을 효과적으로 학습할 수 있기 때문이다. 향후에는 추가적인 데이터 수집과 변수 확장을 통해 모델을 더욱 발전시킬 예정이며, 외부 요인과의 상호작용을 고려하는 새로운 모델의 적용도 검토할 것이다.

이러한 예측 모델은 기업의 에너지 효율성 향상과 경영에 상당한 도움을 줄 것으로 기대된다. 미래에는 추가적인 데이터 수집, 변수 확장, 그리고 모델 튜닝을 통해 성능을 더욱 향상시킬 수 있는 방안을 고려해야 한다. 또한, 새로운 기술과 알고리즘의 도입 및 에너지 소비에 영향을 미치는 다양한 외부 요인을 고려하여 모델을 더욱 발전시킬 수 있을 것이다.

### Reference

- Kim, J., & Hong, J. (2022). "Energy Demand Forecasting and the Relationship between Energy and Economic Growth: A Focus on VECM and ARDL Models." *Applied Economics*, 24(2), 87-113.
- Xiao, L., Shao, W., Liang, T., & Wang, C. (2016). "A Combined Model Based on Multiple Seasonal Patterns and Modified Firefly Algorithm for Electrical Load Forecasting." *Applied Energy*, 167, 135-153. DOI: 10.1016/j.apenergy.2016.01.050
- Simeone, O. (2018). "A Very Brief Introduction to Machine Learning with Applications to Communication Systems." *IEEE Transactions on Cognitive Communications and Networking*, 4(4), 648-664. DOI: 10.1109/TCCN.2018.2881442.
- Sathishkumar, V. E., Shin, C., Cho, Y. (2021). "Efficient Energy Consumption Prediction Model for a Data Analytic-Enabled Industry Building in a Smart City." *Building Research & Information*, 49(1), 127-143.
- De Cauwer, C., Van Mierlo, J., Coosemans, T. (2015). "Energy Consumption Prediction for Electric Vehicles based on Real-world Data." *Energies*, 8(8), 8573-8593.
- Xuemei, L., Yuyan, D., Lixing, D., & Liangzhong, J. (2010). Building Cooling Load Forecasting Using Fuzzy Support Vector Machine and Fuzzy C-mean Clustering. In Proceedings of the 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering, Chengdu, pp. 438-441.
- Zhao, J., Han, Z., Pedrycz, W., & Wang, W. (2015). Granular model of long-term prediction for energy system in steel industry. *IEEE Transactions on Cybernetics*, 46(2), 388-400.
- Dong, B., Cao, C., & Lee, S. E. (2005). Applying Support Vector Machines to Predict Building Energy Consumption in Tropical Region. *Energy Build.*, 37(5), 545-553.
- Lee, S. H. (2003). "A Study on Multicollinearity." Master's thesis, [Yonsei University Graduate School], Seoul.
- Sandels, C., Widén, J., Nordström, L., & Andersson, E. (2015). "Day-ahead Predictions of Electricity Consumption in a Swedish Office Building from Weather, Occupancy, and Temporal Data." *Energy and Buildings*, 108, 279-290. DOI: 10.1016/j.enbuild.2015.08.052.
- Bacher, P., Madsen, H., Nielsen, H. A., & Perers, B. (2013). "Short-term Heat Load Forecasting for Single Family Houses." *Energy and Buildings*, 65, 101-112. DOI: 10.1016/j.enbuild.2013.04.022.