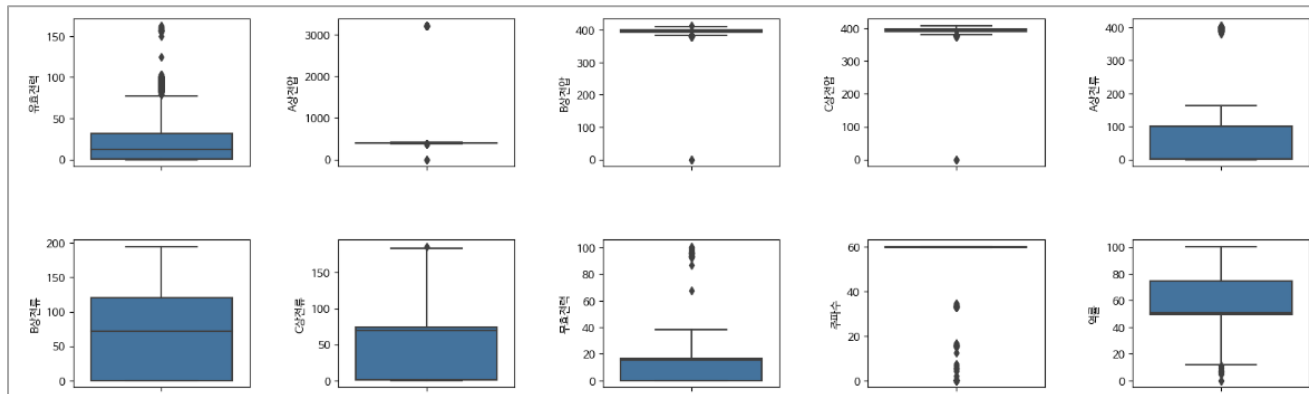


제조 데이터를 기반으로 전력 손실량 예측

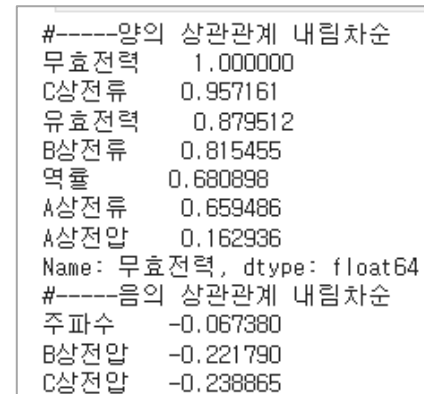
데이터 요약

- 사용될 feature는 10개 (독립변수 : 유효전력, A/B/C상 전압, A/B/C상 전류, 역률, 주파수, 종속변수 : 무효전력), 모두 float

EDA. 데이터 탐색을 통해 데이터 분포 및 이상치 확인 및 상관관계 확인



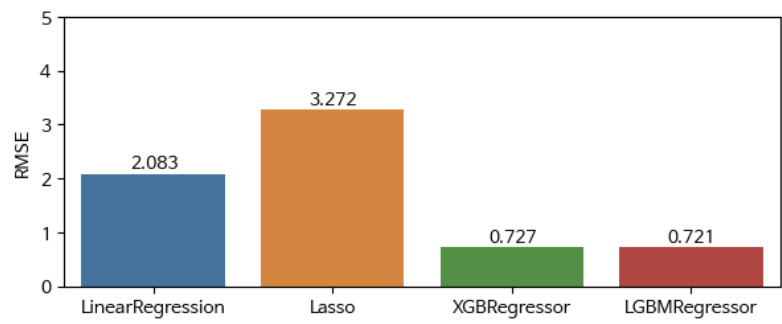
<그림 1. 데이터 분포 및 이상치 확인 시각화>



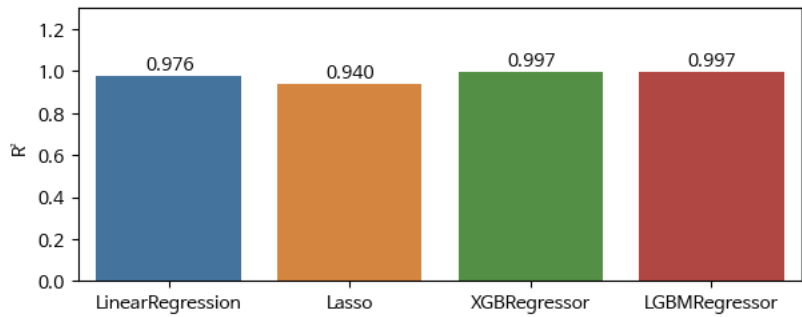
<그림 2. 상관관계수 확인>

1. Boxplot(그림 1)을 사용해 본 결과, 대부분 데이터에 이상치 존재. 종속변수 또한 가지고 있어 삭제하지 않고 진행
2. 종속변수인 무효전력과 상관관계가 높은 'C상전류, 유효전력, B상전류'가 증가하게 되면 '무효전력' 또한 증가할 것으로 가설 설정

■ 모델 성능 평가 및 구축 Linear, Lasso, XGBoost, LightGBM 모델 성능 비교 및 구축



<그림 3. 4가지 회귀모델 RMSE 값 그래프>



<그림 4. 4가지 회귀모델 R² 값 그래프>

- 1. 4가지 회귀모델(Linear, Lasso, XGBoost, LightGBM)으로 성능 평가 했을 때, 정확도와 설명력이 가장 높은 XGBoost로 모델 구축
- 2. K-Fold Cross Validation과 Hyper Parameter(estimatios, learning_rate 조정)를 진행했지만 모두 비슷한 수치를 보여 해당 과정은 영향이 거의 없다는 것을 확인.

■ 결과

	유효전력	A상전압	B상전압	C상전압	A상전류	B상전류	C상전류	주파수	역률
0	79834.1	2920.14	571.33	384.23	248.241	171.23	150.341	60	42.8
1	68343.12	356.123	410.9	382.55	200.44	150.235	180.243	59.75	100.434
2	58500.1	397.2544	398.13	400.231	0.2423	0.0423	0.1332	60	60
3	0	0	0	0	0	0	0	0	0

<그림 5. 가상데이터>

[9.8603630e+01 9.7876518e+01 1.2041444e+01 1.6952663e-03]

<그림 6. 가상데이터 기준 예측 값>

가상의'유효전력, B상전류, C상전류'가 높은 데이터 행의 무효전력 예측 값을 봤을 때, 수치가 높게 나왔으므로 설정한 가설 채택