

2023.11.26

전기 분해를 통한 특정 소재 생산 공정 데이터 분석 보고서

윤혜준

<목차>

1. DATA 소개 및 분석 목적

2. EDA

3. DATA 전처리

4. 모델링 CASE 1

5. 모델링 CASE 2

6. 총정리 및 비즈니스 인사이트

1. <DATA 소개 및 분석 목적>

전기 분해를 통해 특정 소재를 생산하는 공정의 데이터이다.

train 과 test 파일 각각은 동일한 컬럼을 가지고 있으며 다음과 같은 컬럼으로 이루어져 있다.

A. 컬럼소개

i. 관리변수

1. A: 제품 생산량 (값이 클수록 좋음)
2. B: 최종 제품 형태 전환율 (값이 클수록 좋음)
3. C: 전류효율 (값이 클수록 좋음)
4. D: 폐기물 1 농도 (값이 작을수록 좋음)
5. E: 폐기물 2 농도 (값이 작을수록 좋음)

ii. 조업변수

1. 30 개의 조업 변수(ex. 온도,압력,속도 등)

B. 분석목적

- i. 관리 변수 5 개를 토대로 품질 및 생산성 향상 및 원가 절감을 위한 모델링 및 비즈니스 인사이트 도출

2. <EDA>

A. Data 형태

- Train data 형태 : (row 개수 : 43865, column 개수: 36)
- Test data 형태 : (row 개수 : 18798, column 개수: 36)

B. 결측치 여부 확인

- 두 data 모두 결측치 없음

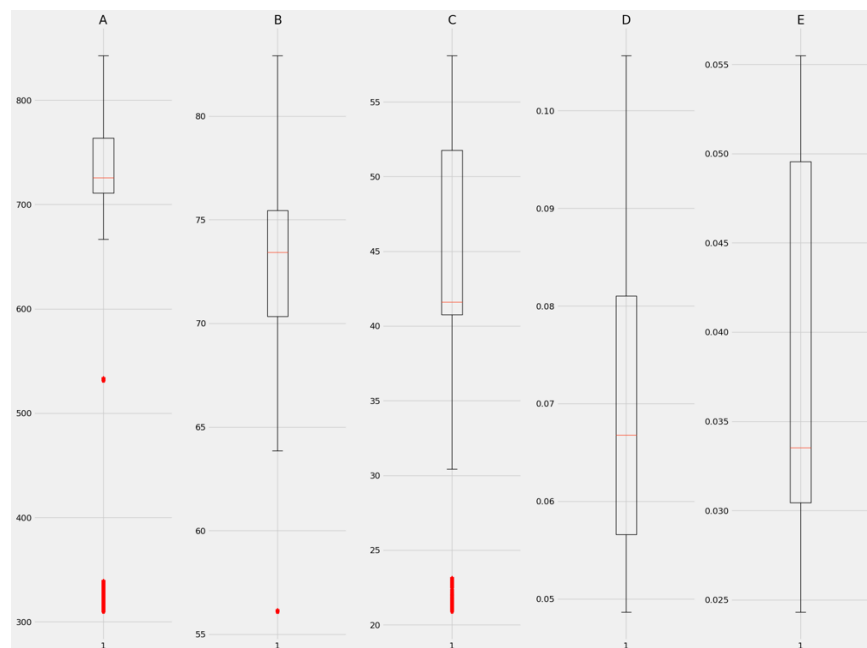
C. Column 정의

- oper_features
 - ✓ X1 ~ X30 :30 개의 조업 변수(ex. 온도,압력,속도 등)
- control_features
 - ✓ A: 제품 생산량 (값이 클수록 좋음)
 - ✓ B: 최종 제품 형태 전환율 (값이 클수록 좋음)
 - ✓ C: 전류효율 (값이 클수록 좋음)
 - ✓ D: 폐기물 1 농도 (값이 작을수록 좋음)
 - ✓ E: 폐기물 2 농도 (값이 작을수록 좋음)

D. 이상치 시각화 및 제거

E. 관리변수 (control_features) 에 대해 이상치 시각화

- Train



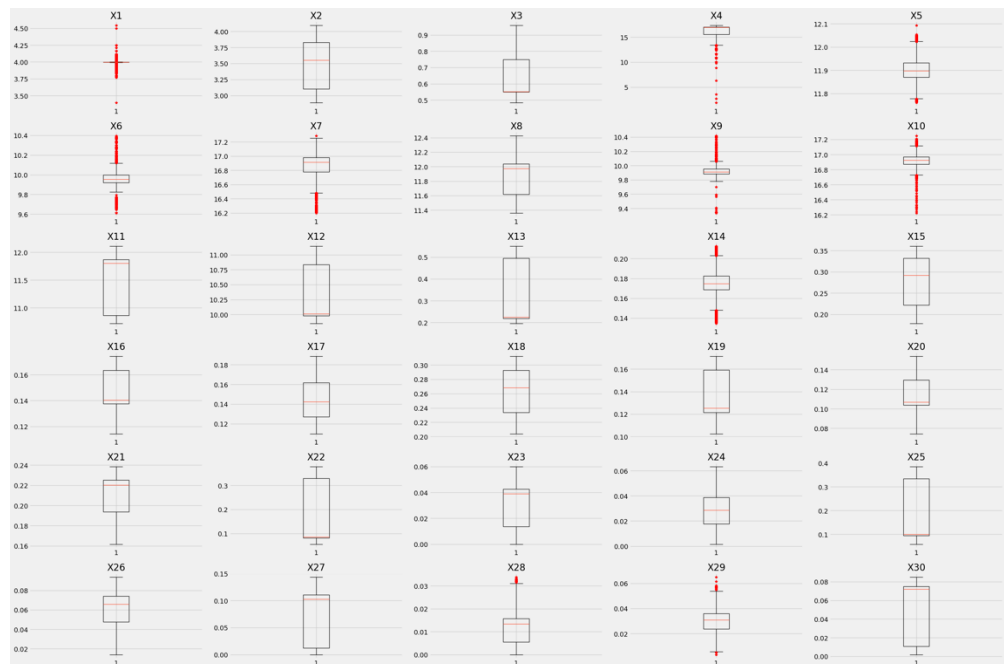
- Test



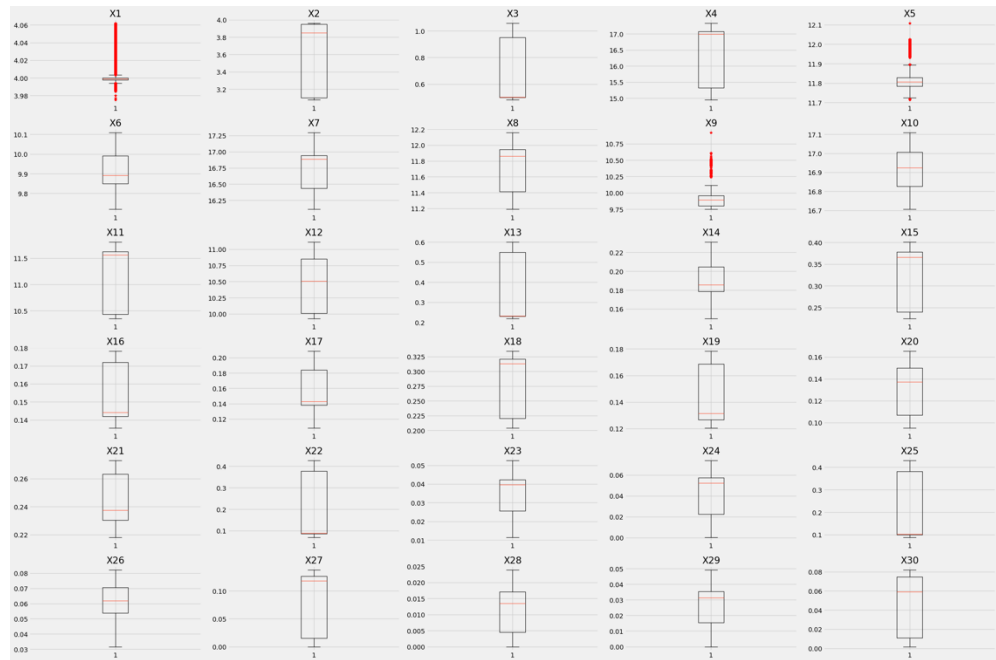
- 이상치 제거 후 데이터 형태 변화 (전 -> 후)
 - ✓ Train data : (43865, 36) -> (43367, 36)
 - ✓ Test data : (18798, 36) -> (17570, 36)

F. 조업변수 (oper_features) 에 대해 이상치 시각화

- Train



- Test



- 이상치 제거 후 데이터 형태 변화 (전 -> 후)

- ✓ Train data : (43865, 36) -> (41006, 32)
- ✓ Test data : (18798, 36) -> (17570, 32)
- ✓ 관리변수 개수 그대로, 조업변수 개수 4 개 감소

G. Feature 별 시각화(각 조업변수와 관리변수 전체 산점도) 후 해석

- ✓ X2 : A 가 750 이 넘으면 X2 가 일정하게 낮은값 , C 가 50 이 넘으면 X2 가 일정하게 낮은 값 , E 가 0.04 이 넘으면 X2 가 일정하게 낮은값
- ✓ X7 : C, E 에 대해서 군집화 된 모습
- ✓ X8 : A, B, C, D, E 애 대해서 군집화 된 모습
- ✓ X10 : C, E 에 대해서 군집화 된 모습
- ✓ X11 : C, E 에 대해서 군집화 된 모습
- ✓ X12 : D0.05-0.06 구간에서 급격히 감소후 일정값, E 에 대해 일정값 보이다가 0.045 부터 급격히 증가
- ✓ X13 : D0.05-0.06 구간에서 급격히 감소후 일정값, E 에 대해 일정값 보이다가 0.045 부터 급격히 증가
- ✓ X15 : C, E 에 대해서 군집화 된 모습, D 와는 비례 추세
- ✓ X16 : A, B, C, D, E 에 대해서 군집화 된 모습

- ✓ X17 : A, B 에 대해 비례상승 추세 , C, D, E 에 대해서 군집화 된 모습
 - ✓ X18 : A, B 에 대해 반비례 추세, D 에 대해 비례 추세, C, E 에 대해서 군집화 된 모습
 - ✓ X19 : A, B, C, D, E 에 대해서 군집화 된 모습
 - ✓ X20 : A, B, C, E 에 대해서 군집화 된 모습, D 에 대해 반비례 추세
 - ✓ X21 : C, E 에 대해서 군집화 된 모습
 - ✓ X22 : A, B, C 에 대해 일정패턴 보임, D0.05-0.06 구간에서 급격히 감소후 일정값, E 에 대해 일정값 보이다가 0.045 부터 급격히 증가
 - ✓ X23 : C, E 에 대해서 군집화 된 모습
 - ✓ X24 : C, E 에 대해서 군집화 된 모습
 - ✓ X25 : A, B, C 에 대해 일정패턴 보임, D0.05-0.06 구간에서 급격히 감소후 일정값, E 에 대해 일정값 보이다가 0.045 부터 급격히 증가
 - ✓ X26 : A, B 에 대해 반비례 추세, C, E 에 대해서 군집화 된 모습
 - ✓ X27 : A, B, C, D, E 에 대해서 군집화 된 모습
 - ✓ X28 : C, E 에 대해서 군집화 된 모습
 - ✓ X29 : C, E 에 대해서 군집화 된 모습
 - ✓ X30 : C, E 에 대해서 군집화 된 모습, D0.05-0.06 구간에서 급격히 증가후 일정값, E 와 반비례 추세
- ⇒ C, E 에 대해서 군집화 된 모습 을 띄는 패턴이 많이 보이고, 구간에서 급격히 감소나 증가후 일정 값을 보이는 패턴이 많이 보임
- ⇒ 비슷한 조업변수의 특징들로 인해 상관관계 분석으로 상관계수 파악

H. 상관계수 시각화 후 해석

0.00-0.19: 매우 약한 상관관계

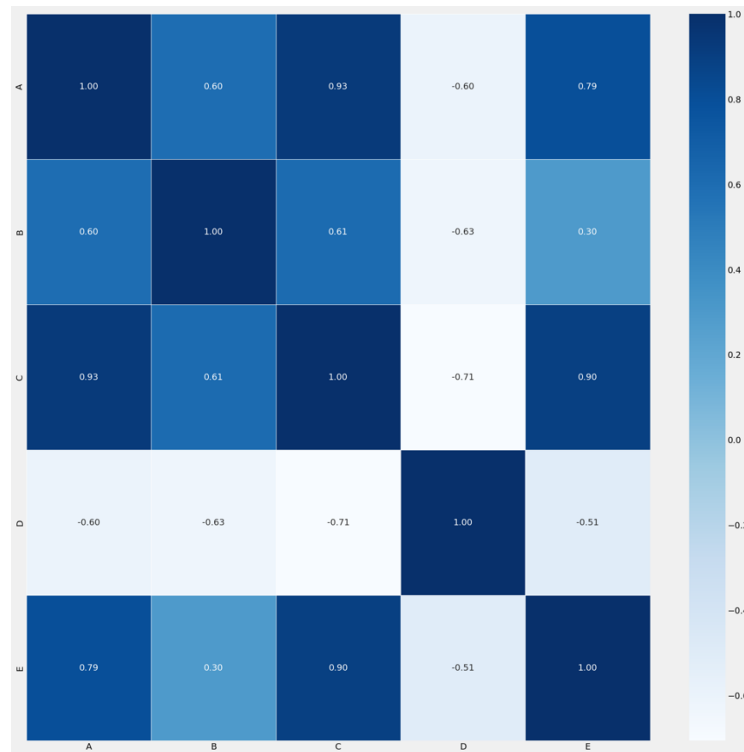
0.20-0.39: 약한 상관관계

0.40-0.59: 중간 정도의 상관관계

0.60-0.79: 강한 상관관계

0.80-1.00: 매우 강한 상관관계

I. 관리변수간 상관계수



① 관리 변수끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.

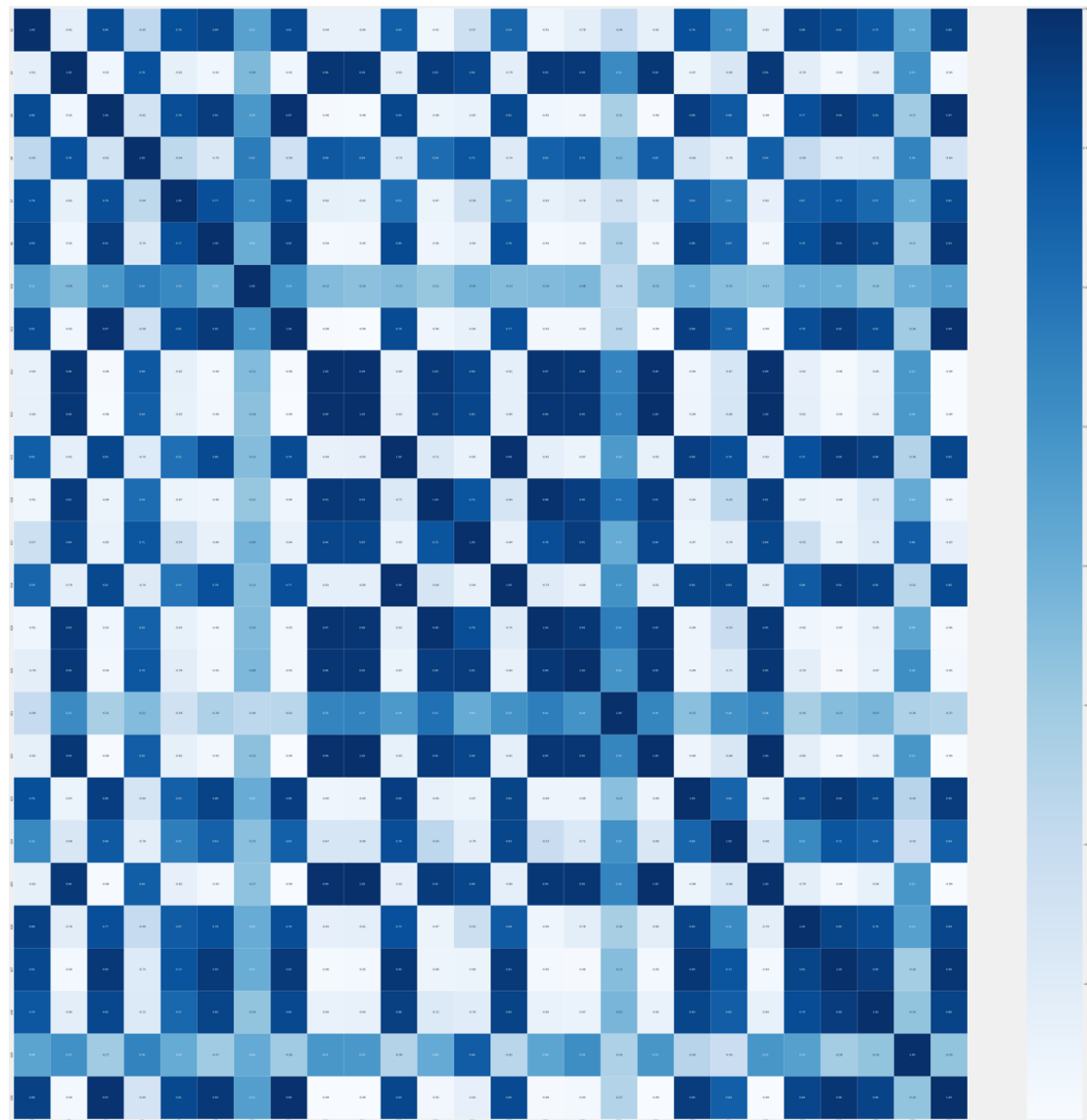
② 가장 높은 상관관계는 A, C 로 0.93

③ 그 다음 가장 높은 상관관계는 C, E 로 0.90

④ 그 다음 높은 상관계수 A, E 로 0.79

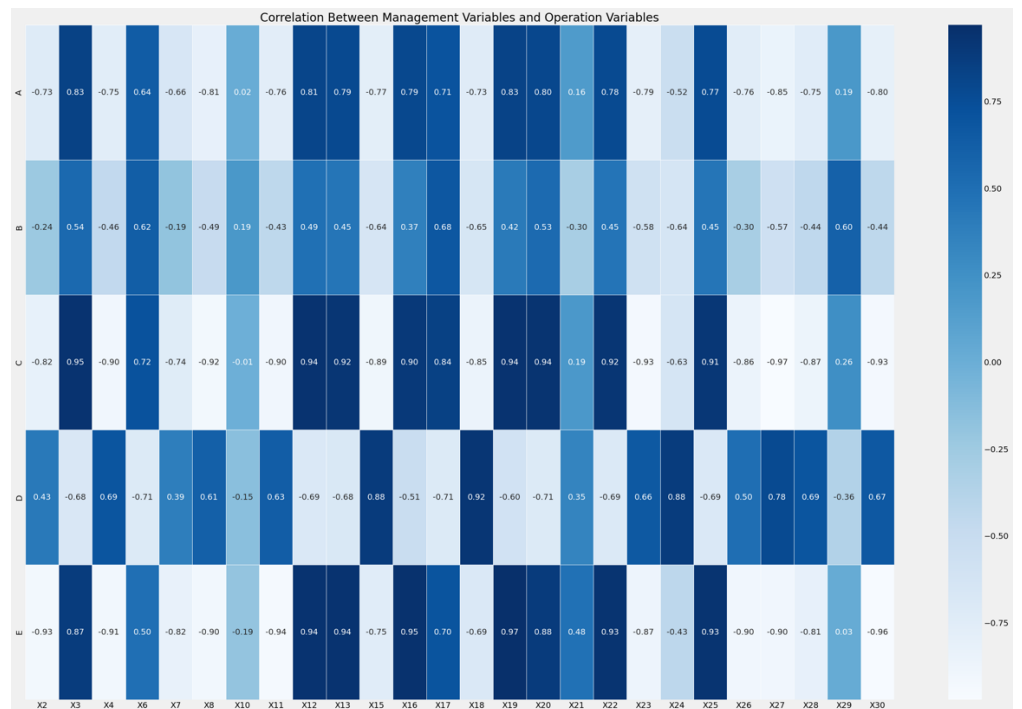
=> A, B, C 는 증가할수록 그리고 D, E 는 작아질수록 품질 및 생산성이 좋아지고 원가도 낮아지지만,
A, C 와 C, E 처럼 매우 강한 양의 상관관계를 갖고있는 변수들로
보아 A, B, C 를 증가시키되 E 의 증가또한 염두해야 한다는 것을
알수 있음.

II. 조업변수간 상관계수



- ① 조업 변수끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.
- ② 예를 들어, 가장 높은 상관관계는 X12, X13 으로 0.99 의 매우 강한 상관성을 띤다.
- ③ 산점도에도 나타났듯이 비슷한 패턴을 띄는 조업변수들끼리 강한 상관성이 나타나고 있다.

III. 조업변수와 관리변수 상관관계 히트맵



- ① 조업 변수와 관리변수 끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.
- ② 가장 높은 상관관계는 E, X19 으로 0.97 의 매우 강한 상관성을 띤다.
==> 특히 A, C, E 는 조업변수와 상관계수가 매우크다

I. 다중공선성 해석

다중공선성 확인 : 일반적으로 VIF 지수 10 이상을 다중공선성 확률 높은 변수로 취급

i. 조업변수 다중공선성 확인

Variable	VIF
0 X2	1.939805e+03
1 X3	1.239772e+03
2 X4	1.852053e+04
3 X5	5.778393e+05
4 X6	5.053102e+05
5 X7	1.051342e+05
6 X8	7.864117e+04
7 X9	4.358935e+05
8 X10	3.708410e+05
9 X11	2.345937e+05
10 X12	1.419204e+05
11 X13	inf
12 X15	1.385593e+04
13 X16	inf
14 X17	inf
15 X18	inf
16 X19	inf
17 X20	inf
18 X21	inf
19 X22	inf
20 X23	2.249446e+03
21 X24	1.552608e+02
22 X25	inf
23 X26	7.314531e+03
24 X27	1.809719e+03
25 X28	1.791158e+04
26 X29	inf
27 X30	inf

ii. 관리변수에 대해 다중공선성 확인

Variable	VIF
0 A	1752.134147
1 B	926.875587
2 C	1489.685921
3 D	62.356247
4 E	183.172258

J. EDA 결론

1. 관리 변수끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.
 - a. 가장 높은 상관관계는 A, C 로 0.93
 - b. 그다음 가장 높은 상관관계는 C, E 로 0.90
 - c. 그다음 높은 상관관계수 A, E 로 0.79
 - d. A, B, C 는 증가할수록 그리고 D, E 는 작아질수록 품질 및 생산성이 좋아지고 원가도 낮아지지만 A, C 와 C, E 처럼 매우 강한 양의 상관관계를 갖고있는 변수들로 보아 A, B, C 를 증가시키면 E 의 증가또한 염두해야 한다는 것을 알수 있음.
2. 조업 변수끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.
 - a. 가장 높은 상관관계는 X12, X13 으로 0.99 의 매우 강한 상관성을 띈다.
 - b. 산점도에도 나타났듯이 비슷한 패턴을 띄는 조업변수들끼리 강한 상관성이 나타나고 있다.
3. 조업 변수와 관리변수 끼리 대부분이 중간이상의 양/음의 상관관계를 갖고있다.
 - a. 가장 높은 상관관계는 E 와 X19 으로 0.97 의 매우 강한 상관성을 띈다. ==> 특히 A, C, E 는 조업변수와 상관관계수가 매우크다.
4. 변수들끼리의 다중공선성이 나타나고, 중요도를 파악하여 인사이트를 얻기 위해 때문에 앙상블 모델링 진행.
5. train, test data 의 outlier 에 대해 처리를 진행했지만 분포가 다른 변수들이 많은것으로 보아 train 에 대한 model 의 과적합 예상된다.
6. 변수별 중요도를 앙상블 모델들에서 추출 후 가장 많이 나타나는 변수들이 관리변수를 최적화 하는데 중요한 요인으로 파악될것으로 보인다.
7. Case1(개별 관리변수에 대한 모델링)과 Case2(관리변수 통합하여 새로운 관리변수에 대한 모델링)를 나누어 진행하여

개별 관리변수에 대한 모델의 설명력 뿐만 아니라, 통합 관리변수 모델링의 설명력 또한 확인한다.

3. <DATA 전처리>

- A. Train 데이터에 대해 8:2 로 train, val 데이터 split
- B. Train, Test 데이터에 대해 작업변수 X, 관리변수 y 로 설정
- C. Normalization 진행 (평균 0 , 표준편차 1)
 - i. Train 데이터 X 에 대해 표준화를 적용하고 Test 데이터의 X 에 대해서도 표준화를 적용
- D. Y(관리변수)변수 변형시키기
 - i. 모델링 2 를 위해서 관리변수들을 normalization 후 new_y 로 변형
 - ii. 각 관리 변수의 중요도가 같다고 가정
 - iii. $new_y = A + B + C - D - E$

4. <모델링 CASE 1>

- A. 관리변수 (A, B, C, D, E)를 각각 결과변수로 모델링 진행
ex) $y_train_A, y_val_A = y_train['A'], y_val['A']$
- B. Light GBM Regression 모델 사용
LightGBM Regression 은 그래디언트 부스팅 알고리즘을 활용하여 회귀 문제를 해결하는 모델이며 여러 결정 트리를 순차적으로 학습하여 예측 성능을 향상시킨다.
 - i. Validation data 에 대한 성능

	Variable	MAE	R ²	MSE	RMSE
0	A	4.764590	0.964351	4.633970e+01	6.807327
1	B	0.588307	0.943390	7.331919e-01	0.856266
2	C	0.241708	0.996382	1.297614e-01	0.360224
3	D	0.000407	0.998436	3.455899e-07	0.000588
4	E	0.000133	0.999580	3.841330e-08	0.000196

- ii. Test data 에 대한 성능

	Variable	MAE	R ²	MSE	RMSE
0	A	19.552048	0.580312	579.978698	24.082747
1	B	2.420380	0.284429	8.109953	2.847798
2	C	1.151182	0.953130	1.958842	1.399586
3	D	0.004933	0.925508	0.000038	0.006191
4	E	0.001588	0.939718	0.000004	0.002022

C. RandomForest 모델 사용

i. Validation data 에 대한 성능

	Variable	MAE	R ²	MSE	RMSE
0	A	2.964766	0.980578	2.524589e+01	5.024529
1	B	0.358176	0.968011	4.143099e-01	0.643669
2	C	0.198104	0.997010	1.072452e-01	0.327483
3	D	0.000306	0.998788	2.678592e-07	0.000518
4	E	0.000094	0.999714	2.616984e-08	0.000162

ii. Test data 에 대한 성능

	Variable	MAE	R ²	MSE	RMSE
0	A	22.640974	0.432068	784.841132	28.015016
1	B	2.554663	0.182210	9.268453	3.044413
2	C	2.244832	0.737615	10.965871	3.311476
3	D	0.004754	0.927529	0.000037	0.006107
4	E	0.001925	0.910875	0.000006	0.002458

■ LGBM, RandomForest 성능비교

◆ 사용한 지표 :

MAE : 실제 값과 예측 값 간의 절대적인 차이를 평균한 값

R-squared : R²는 회귀 모델의 설명력을 나타내는 지표로서 종속 변수의 총 변동 중 모델에 의해 설명되는 비율이다. R² 값이 1 에 가까울수록 모델이 데이터를 잘 설명한다고 볼 수 있다.

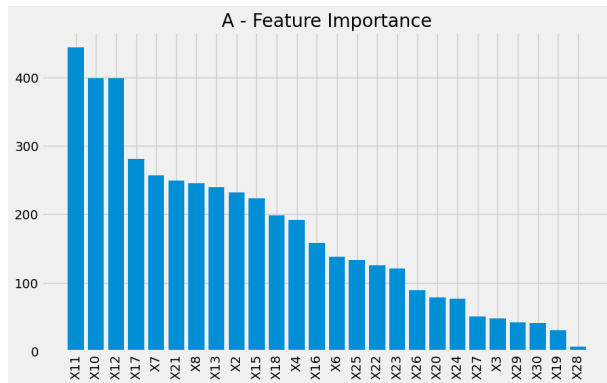
MSE : 평균 제곱 오차를 나타내는 지표로, 실제 값과 예측 값 간의 제공된 차이를 평균한 값, 값이 작을수록 모델의 예측이 정확하다고 평가

R-squared : R²는 회귀 모델의 설명력을 나타내는 지표로서 종속 변수의 총 변동 중 모델에 의해 설명되는 비율이다. R² 값이 1 에 가까울수록 모델이 데이터를 잘 설명한다고 볼 수 있다.

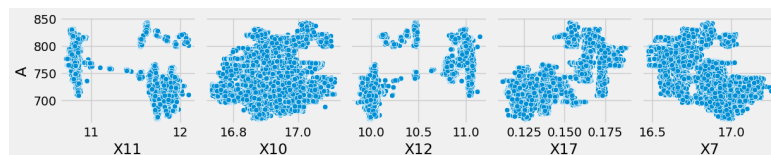
- ◆ test data 에 대하여 두가지모델을 MAE, R^2 , MSE, RMSE 를 지표로 사용하여 비교했을 때, 전반적으로 LGBM 이 모델을 더 잘 설명하고 있음을 알 수 있다.
- ◆ 특히 C, D, E 에 대해서는 설명력(R-squared)도 매우 높으며 예측성능 또한 매우 높다.
- ◆ 더 성능이 높은 LGBM 모델에 대해서 중요성 높은 조업변수 찾는 과정 진행

D. LGBM 모델에서 중요도 높은 조업변수 탐색

i. 관리변수 A 를 결과변수로 사용한 모델



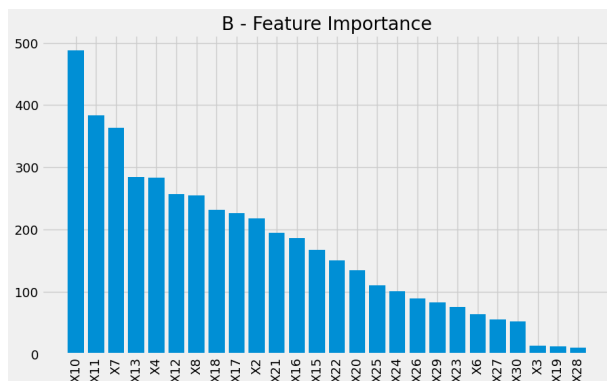
1. 중요도가 가장 높은 변수 5 개 선택 후 산포도 시각화



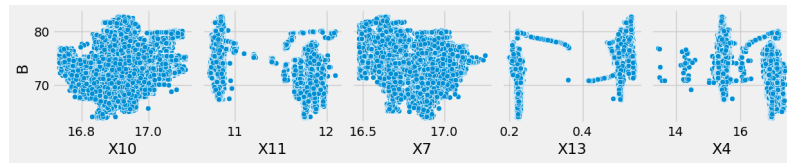
2. 선택된 특성들과 A 변수 간의 상관 계수 계산

X11	X10	X12	X17	X7
-0.762789	0.021186	0.813873	0.707558	-0.653501

ii. 관리변수 B 를 결과변수로 사용한 모델



1. 중요도가 가장 높은 변수 5 개 선택 후 산포도 시각화

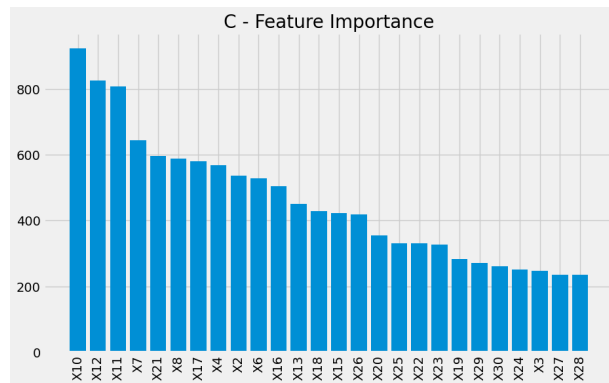


2. 선택된 특성들과 A 변수 간의 상관 계수 계산

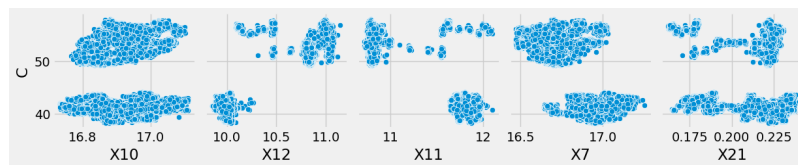
X10 X11 X7 X13 X4

0.192683 -0.436761 -0.189201 0.452635 -0.460702

iii. 관리변수 C 를 결과변수로 사용한 모델



1. 중요도가 가장 높은 변수 5 개 선택 후 산포도 시각화

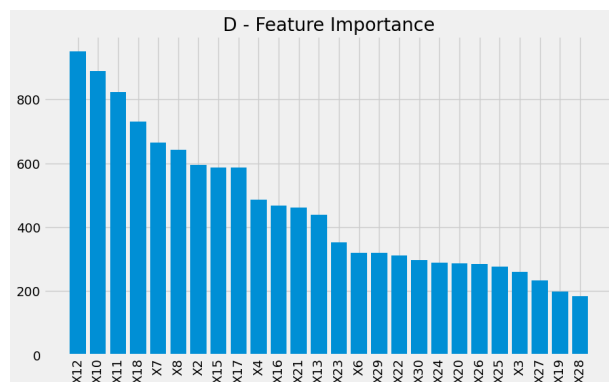


2. 선택된 특성들과 A 변수 간의 상관 계수 계산

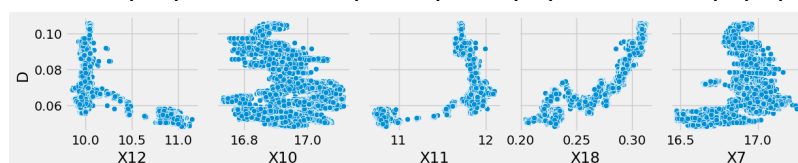
X10 X12 X11 X7 X21

-0.012691 0.941277 -0.899078 -0.734221 0.183341

iv. 관리변수 D 를 결과변수로 사용한 모델



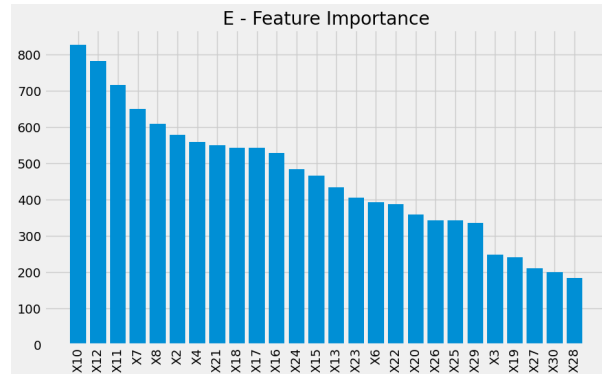
1. 중요도가 가장 높은 변수 5 개 선택 후 산포도 시각화



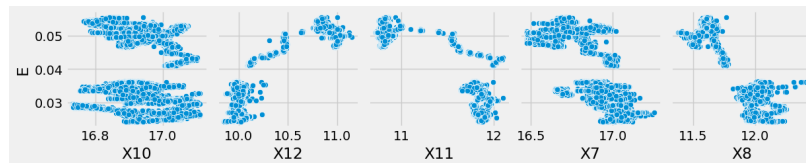
2. 선택된 특성들과 A 변수 간의 상관 계수 계산

X12	X10	X11	X18	X7
-0.689538	-0.146162	0.631502	0.917352	0.390517

v. 관리변수 E 를 결과변수로 사용한 모델



1. 중요도가 가장 높은 변수 5 개 선택 후 산포도 시각화



2. 선택된 특성들과 A 변수 간의 상관 계수 계산

X10	X12	X11	X7	X8
-0.189024	0.939882	-0.936909	-0.818233	-0.903058

E. 모델링 CASE 1 - Insight 정리

i. A, B, C 에 대하여

- A, B, C 는 클수록 공정에서 좋은 결과를 가져오는 관리변수이기때문에, 양의상관관계를 갖고있는 변수들을 증가시키고, 음의 상관관계를 갖는 변수들을 감소시킬 필요가 있다.
- A 와 각 변수에 대한 상관계수는
X11, X10, X12, X17, X7 에 대하여
-0.762789 0.021186 0.813873 0.707558 -0.653501 이고,
따라서 X11, X7 을 감소시키고, X12, X17 를 증가하는 방향으로
변화 시키면 A 값이 커지는 방향으로 변화될것으로 예상할 수 있다.
- B 와 각 변수에 대한 상관계수는
X10, X11, X7, X13, X4 에 대하여
0.192683 -0.436761 -0.189201 0.452635 -0.460702 이고,

따라서 X11, X7, X4 을 감소시키고, X10, X13 를 증가하는 방향으로 변화 시키면 B 값이 커지는 방향으로 변화될것으로 예상할 수 있다.

4. C 와 각 변수에 대한 상관계수는 X10, X12, X11, X7, X21 에 대하여

-0.012691 0.941277 -0.899078 -0.734221 0.183341 이고,

따라서 X11, X7 을 감소시키고, X12, X21 를 증가하는 방향으로 변화 시키면 C 값이 커지는 방향으로 변화될것으로 예상할 수 있다.

ii. D, E 에 대하여

1. D, E 는 작을수록 공정에서 좋은 결과를 가져오는 관리변수이기때문에, 음의 상관관계를 갖고있는 변수들을 증가시키고, 양의 상관관계를 갖는 변수들을 감소시킬 필요가 있다.

2. D 와 각 변수에 대한 상관계수는

X12, X10, X11, X18, X7 에 대하여

-0.689538 -0.146162 0.631502 0.917352 0.390517 이고,

따라서 X11, X18, X7 을 감소시키고, X12, X10 를 증가하는 방향으로 변화 시키면 D 값이 작아지는 방향으로 변화될것으로 예상할 수 있다.

3. E 와 각 변수에 대한 상관계수는 X10, X12, X11, X7, X8 에 대하여

-0.189024 0.939882 -0.936909 -0.818233 -0.903058 이고,

따라서 X12 을 감소시키고, X10, X11, X7, X8 를 증가하는 방향으로 변화 시키면 E 값이 커지는 방향으로 변화될것으로 예상할 수 있다.

5. <모델링 CASE 2>

A. new_Y ($A + B + C - D - E$)에 대해 모델링 진행

B. Light GBM Regression 모델 사용

LightGBM Regression 은 그래디언트 부스팅 알고리즘을 활용하여 회귀 문제를 해결하는 모델이며 여러 결정 트리를 순차적으로 학습하여 예측 성능을 향상시킨다.

i. Validation data 에 대한 성능

```
validation 세트 MAE: 0.2377093239186032
validation 세트 R²: 0.9842118013902373
validation 세트 RMSE: 0.3553576640208286
```

ii. Test data 에 대한 성능

```
테스트 세트 MSE: 3.484575256526753
테스트 세트 MAE: 1.5794581675392463
테스트 세트 R²: 0.7016897386020507
테스트 세트 RMSE: 1.8667017052884356
```

C. RandomForest 모델 사용

i. Validation data 에 대한 성능

```
validation 세트 MSE: 0.20753568590174884
validation 세트 MAE: 0.3004729357285662
validation 세트 R²: 0.9743941955780562
validation 세트 RMSE: 0.45255212429554237
```

ii. Test data 에 대한 성능

```
테스트 세트 MSE: 3.7683906543547643
테스트 세트 MAE: 1.6703003598000503
테스트 세트 R²: 0.6773926466232634
테스트 세트 RMSE: 1.941234312069196
```

■ LGBM, RandomForest 성능비교

◆ 사용한 지표 :

MAE : 실제 값과 예측 값 간의 절대적인 차이를 평균한 값

R-squared : R^2 는 회귀 모델의 설명력을 나타내는 지표로서 종속 변수의 총 변동 중 모델에 의해 설명되는 비율이다. R^2 값이 1 에 가까울수록 모델이 데이터를 잘 설명한다고 볼 수 있다.

MSE : 평균 제곱 오차를 나타내는 지표로, 실제 값과 예측 값 간의 제공된 차이를 평균한 값, 값이 작을수록 모델의 예측이 정확하다고 평가

R-squared : R^2 는 회귀 모델의 설명력을 나타내는 지표로서 종속 변수의 총 변동 중 모델에 의해 설명되는 비율이다. R^2 값이 1에 가까울수록 모델이 데이터를 잘 설명한다고 볼 수 있다.

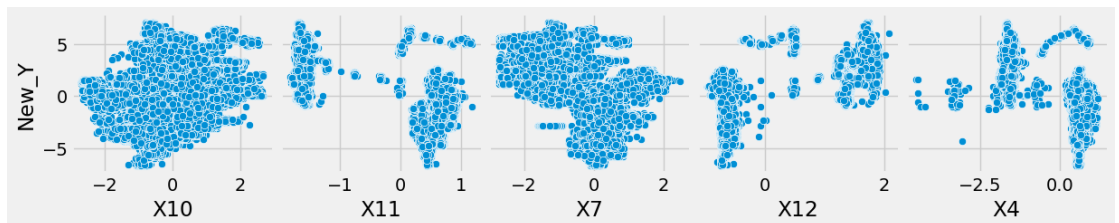
- ◆ test data에 대하여 두가지모델을 MAE, R^2 , MSE, RMSE를 지표로 사용하여 비교했을 때, 전반적으로 LGBM 이 모델을 더 잘 설명하고 있음을 알 수 있다. 특히 모델의 설명력이 70%이상으로 매우 높다.
- ◆ 더 성능이 높은 LGBM 모델에 대해서 중요성 높은 조업변수 찾는 과정 진행

D. LGBM 모델에서 중요도 높은 조업변수 탐색

i. new_y에 대한 LGBM 모델내의 조업변수 중요도

```
Model 3 - Feature Importances:
X10: 986
X11: 869
X7: 766
X12: 758
X4: 615
X8: 594
X17: 584
X2: 562
X21: 530
X16: 522
X15: 483
X13: 445
X18: 435
X6: 411
X26: 350
X20: 334
X22: 316
X25: 302
X27: 299
X24: 299
X23: 296
X29: 282
X3: 267
X30: 245
X19: 216
X28: 186
```

ii. 중요도가 가장 높은 변수 5개 선택 및 산포도 시각화



iii. 새로운 y와 중요도가 높은 변수들 간의 상관계수 계산

	X10	X11	X7	X12	X4
Correlation	0.187824	-0.627949	-0.402429	0.698011	-0.661533

E. 모델링 CASE 2 - Insight 정리

- i. 새로 조합한 $\text{new_y} = A + B + C - D - E$ 는 클수록 공정에서 좋은 결과를 가져오는 관리변수이기때문에, 양의상관관계를 갖고있는 변수들을 증가시키고, 음의 상관관계를 갖는 변수들을 감소시킬 필요가 있다.
- ii. new_y 는 X10, X11, X7, X12, X4 에 대하여 다음과 같은 상관계수를 보인다.
0.187824 -0.627949 -0.402429 0.698011 -0.661533
- iii. - 따라서, new_y 와 양의 상관계수를 갖고있는 X10, X12 를 증가시키고, 음의상관계수를 갖고있는 X11, X7, X4 를 감소하는 방향으로 변화시킨다면 new_y 를증가하는 방향으로 변화시킬 수 있을것이다.

6. <총정리 및 비즈니스 인사이트>

A. 모델링 case 1 – 관리변수별 모델링

- i. A (제품 생산량) 의 값을 증가시키기 위해
: X11, X7 을 감소시키고, X12, X17 를 증가시킨다.
- ii. B (최종 제품 형태 전환율) 의 값을 증가시키기 위해
: X11, X7, X4 을 감소시키고, X10, X13 를 증가시킨다.
- iii. C (전류효율) 의 값을 증가시키기 위해
: X11, X7 을 감소시키고, X12, X21 를 증가시킨다.
- iv. D (폐기물 1 농도) 의 값을 감소시키기 위해
: X11, X18, X7 을 감소시키고, X12, X10 를 증가시킨다.
- v. E (폐기물 2 농도) 의 값을 감소시키기 위해
: X12 을 감소시키고, X10, X11, X7, X8 를 증가시킨다.

관리 변수	감소시킬 조업 변수	증가시킬 조업 변수
A(제품 생산량)	X11, X7	X12, X17
B(최종 제품 형태 전환율)	X11, X7, X4	X10, X13
C(전류효율)	X11, X7	X12, X21
D(폐기물 1 농도)	X11, X18, X7	X12, X10
E(폐기물 2 농도)	X12	X10, X11, X7, X8

B. 모델링 case 2 - 관리변수 통합 후 모델링

- i. $\text{New_y} = A+B+C-D-E$
- ii. X11, X7, X4 를 감소시키고, X12, X10 을 증가시킨다.

관리 변수	감소시킬 조업 변수	증가시킬 조업 변수
New_y	X11, X7, X4	X12, X10

C. 최종 Insight

- i. case 1 의 결과를 종합해 보면, case2 가 각 관리변수들 간의 상호작용을 고려한 통합적인 모델이며 설명력이 뛰어나고 신뢰도가 높은 모델임을 알 수 있다.
- ii. 개별 관리변수의 관리를 위해서는 case1 의 모델의 insight 를 활용하여 A,B,C 의 값을 크게 만들고 D, E 의 값을 작게 만들기 위해 조업변수들을 조절할 수 있을 것 이다.
- iii. 관리 변수 5 개는 밀접한 상호작용을 갖고 있는 변수들이기 때문에 case 2 의 를 사용하여 전반적인 품질향상과 생산성향상에 기여할 수 있을 것이며, 원가 절감에도 기여할 수 있을 것이다.
- iv. 관리 변수 E 의 경우 중요도 높은 조업변수의 경향이 나머지 관리 변수와 다른 것이 존재하기 때문에 이점을 유의하여 조업변수의 변화가 필요하다.