

A photograph of an industrial facility, likely a steel mill, at night. The scene is illuminated by bright orange and yellow lights from the facility's structures and equipment. A large, multi-story building with a red roof is prominent in the foreground. In the background, a tall, illuminated tower or chimney rises into the dark sky. The overall atmosphere is industrial and intense.

# Scale 불량 원인 분석을 통한 제안 제시

B 반 1 조 장 선 영

# 과제 정의

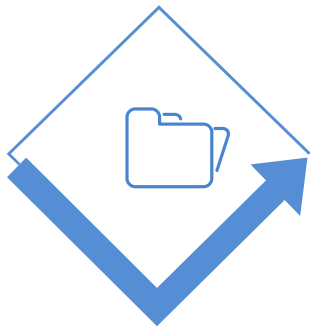
## 분석 배경

OO 공장의 고객센터에서 최근 “Scale 불량 발생 증가”라는 이슈가 발생

원인 분석 결과 압연공정에서 Scale 불량 급증을 확인할 수 있었다.

그래서 데이터를 수집하여 다양한 분석을 통해 불량 발생의 근본 원인을 찾고 결과를 해석하여 개선 기회를 도출하고자 한다.

## Process



### 1. 데이터 구성

### 2. 데이터 품질 확인

- 결측치 확인 및 처리
- 이상치 확인 및 처리



### 3. 시각화

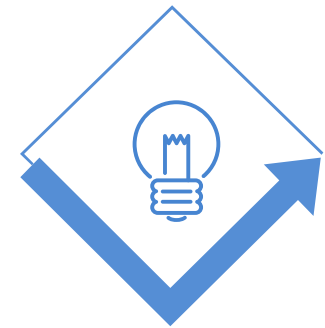


### 5. 모델링

- 다중회귀분석
- 의사결정나무
- 랜덤포레스트
- 그래디언트 부스팅
- SVM



### 6. 모델 평가



### 7. 결론

# 데이터 구성 및 품질 확인

## [배경]

- OO공장의 "Scale 불량 발생 증가" issue 발생
- 원인 분석 결과 압연공정의 Scale 불량 급증 확인

## [주제]

- 불량 발생의 근본 원인 모색
- 근본 원인을 근거로 개선 기회 도출

## 1. 데이터 구성

- **기간:** 2008년 8월 1일 ~ 2008년 8월 2일 (총 720개 data)
- **목표변수:** SCALE (양품/불량)
- **설명변수:** SPEC, STEEL\_KIND, PT\_THICK, PT\_WIDTH, PT\_LENGTH, PT\_WEIGHT, FUR\_NO, FUR\_NO\_ROW, FUR\_HZ\_TEMP, FUR\_HZ\_TIME, FUR\_SZ\_TEMP, FUR\_SZ\_TIME, FUR\_TIME, FUR\_EXTEMP, ROLLING\_TEMP\_T5, HSB, ROLLING\_DESCALING, WORK\_GR

## 2. 데이터 품질 확인

### • 결측치

결측치 존재하지 않음

### • 이상치

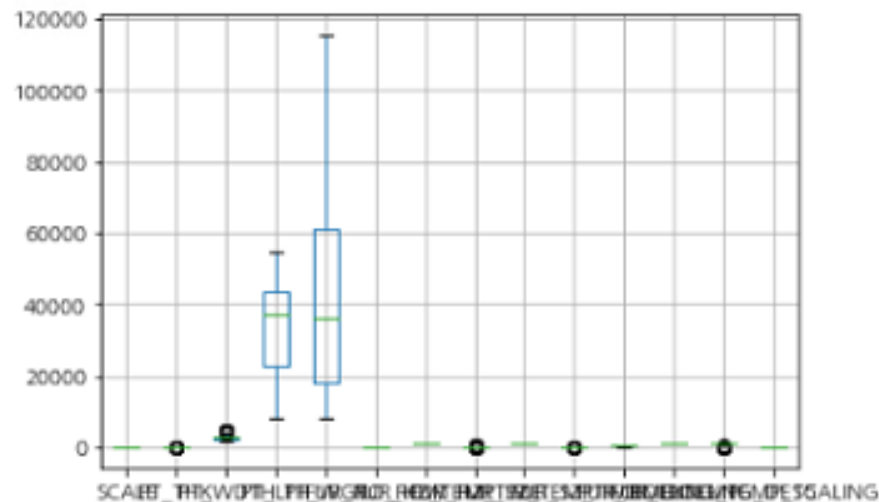
사용자 정의 함수로 각 변수별 IQR방식으로 이상치를 탐지

But,

따로 이상치에 대한 처리는 하지 않음

```
SCALE 0
STEEL_KIND 0
PT_THK 0
PT_WDTH 0
PT_LTH 0
PT_WGT 0
FUR_NO 0
FUR_NO_ROW 0
FUR_HZ_TEMP 0
FUR_HZ_TIME 0
FUR_SZ_TEMP 0
FUR_SZ_TIME 0
FUR_TIME 0
FUR_EXTEMP 0
ROLLING_TEMP_T5 0
HSB 0
ROLLING_DESCALING 0
WORK_GR 0
dtype: int64
```

<결측치>



<이상치> - boxplot

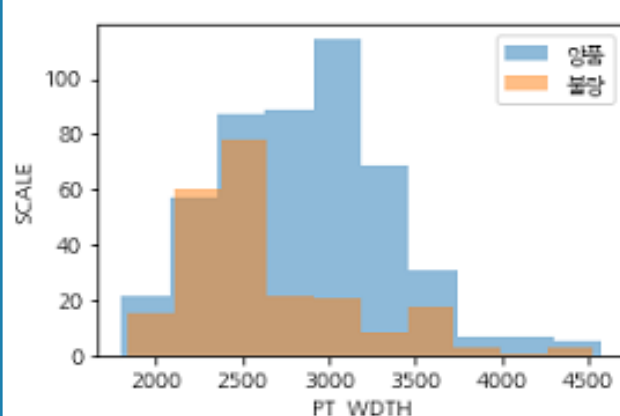
# 시각화

## 설명변수 - 연속형 변수

### Histogram

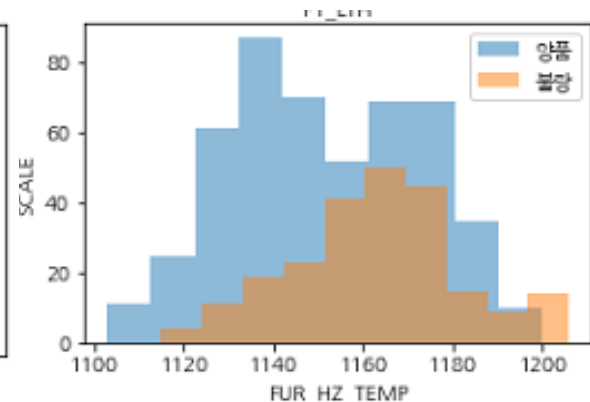
→ 연속형 변수들만 추출해서 Histogram으로 표현

: 변수들 간 상관도를 파악



→ Plate 폭이 작을수록

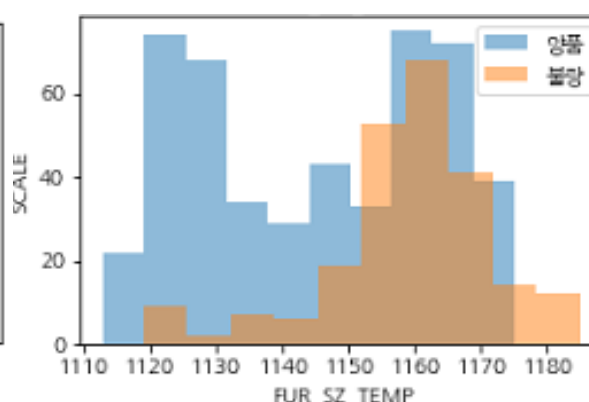
: 불량 경향↑



→ 가열로 가열대 온도가

높을수록

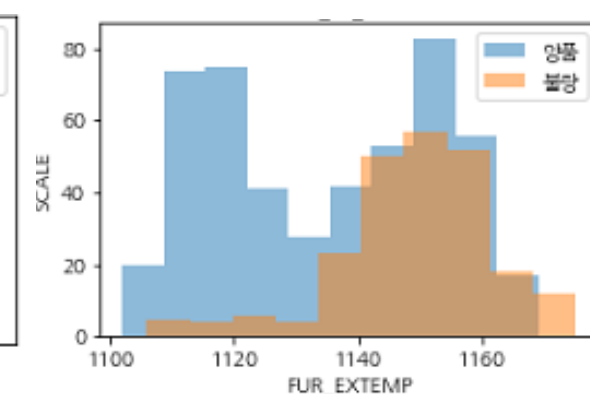
: 불량 경향↑



→ 가열로 균열대 온도가

높을수록

: 불량 경향↑



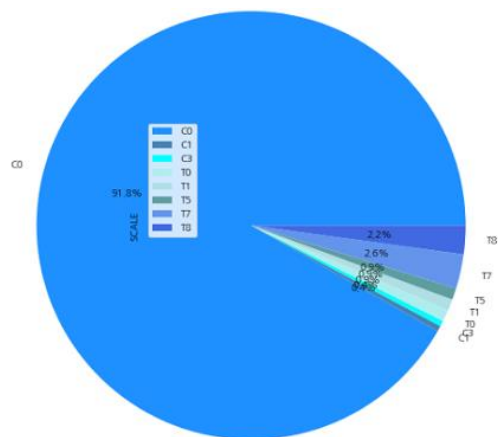
→ 추출 온도가 높을수록

: 불량 경향↑

# 시각화

## 설명변수 – 주요 변수별 Scale과 비교

### 1) 강종

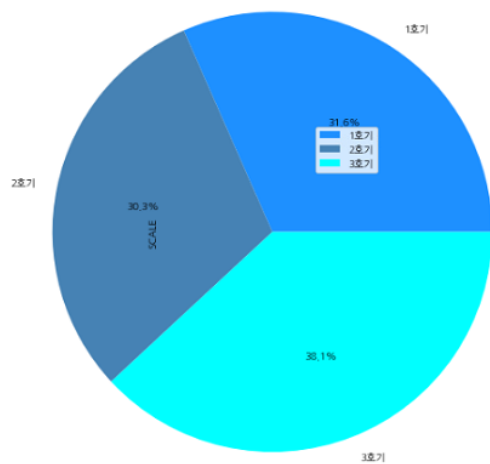


STEEL_KIND	SCALE_PROB
0	C0 42.147117
1	C1 100.000000
2	C3 14.285714
3	T0 12.500000
4	T1 11.111111
5	T3 0.000000
6	T5 4.651163
7	T7 17.142857
8	T8 5.263158

→ C0의 비중이 높음

→ 전체 C0 중 42.1% 불량

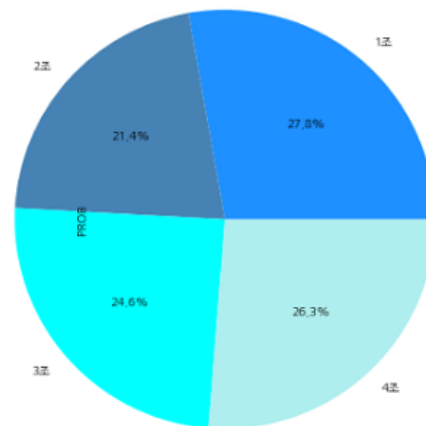
### 2) 가열호



FUR_NO	PROB
0	1호기 0.304
1	2호기 0.295
2	3호기 0.362

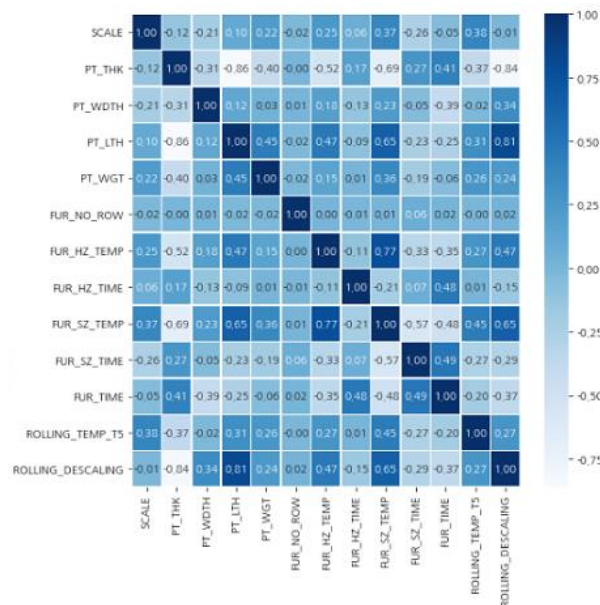
→ 가열호 1, 2, 3호에서 발생하는 불량률 빈도가 유사

### 3) 작업조



WORK_GR	PROB
0	1조 0.354497
1	2조 0.272727
2	3조 0.313953
3	4조 0.335052

→ 1조, 4조 인 경우 상대적으로 불량률이 높을 확률 높음



<참고>

Heatmap을 통한 상관성

→ FUR\_SZ\_TEMP(가열로

균열대 온도)와

FUR\_EXTEMP(압연온도)

: 거의 일치하는 상관성



# 모델링

## (1) 로지스틱회귀분석

Warning: Maximum number of iterations has been exceeded.  
Current function value: 0.360586  
Iterations: 35  
Function evaluations: 36  
Gradient evaluations: 36

Logit Regression Results

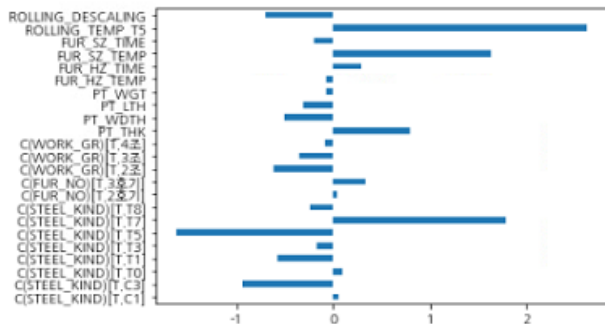
Dep. Variable:	SCALE	No. Observations:	720
Model:	Logit	DF Residuals:	696
Method:	MLE	Pseudo R-squ.:	0.4254
Date:	Wed, 10 Nov 2021	Log-Likelihood:	-259.62
Time:	02:00:03	LL-Null:	-451.80
converged:	False	LLR p-value:	2.930e-67
Covariance Type:	nonrobust		

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.3874	0.443	-3.129	0.002	-2.256	-0.518
C(STEEL_KIND)[T.C1]	0.0554	3.547	0.016	0.988	-6.896	7.007
C(STEEL_KIND)[T.C3]	-0.9428	1.012	-0.932	0.351	-2.926	1.040
C(STEEL_KIND)[T.T0]	0.0972	1.313	0.074	0.941	-2.476	2.671
C(STEEL_KIND)[T.T1]	-0.5771	1.133	-0.509	0.611	-2.798	1.644
C(STEEL_KIND)[T.T3]	-0.1645	3.179	-0.052	0.959	-6.396	6.067
C(STEEL_KIND)[T.T5]	-1.6235	1.431	-1.134	0.257	-4.429	1.182
C(STEEL_KIND)[T.T7]	-1.7911	1.277	-1.403	0.161	-0.712	4.294
C(STEEL_KIND)[T.T8]	-0.2394	1.253	-0.191	0.849	-2.696	2.217
C(FUR_NO)[T.2&27]	0.0465	0.288	0.162	0.872	-0.517	0.610
C(FUR_NO)[T.3&27]	0.3312	0.285	1.164	0.244	-0.227	0.889
C(WORK_GR)[T.2&]	-0.6214	0.323	-1.924	0.054	-1.254	-0.011
C(WORK_GR)[T.3&]	-0.3502	0.331	-1.060	0.289	-0.998	0.298
C(WORK_GR)[T.4&]	-0.0854	0.307	-0.279	0.781	-0.686	0.515
PT_THK	0.7924	0.456	1.736	0.083	-0.102	1.687
PT_WDTH	-0.5028	0.157	-3.205	0.001	-0.810	-0.195
PT_LTH	-0.3048	0.309	-0.987	0.324	-0.910	0.301
PT_WGT	-0.0643	0.135	-0.476	0.634	-0.329	0.200
FUR_HZ_TEMP	-0.0698	0.257	-0.271	0.786	-0.574	0.434
FUR_HZ_TIME	0.2882	0.130	2.212	0.027	0.033	0.544
FUR_SZ_TEMP	1.6407	0.452	3.627	0.000	0.754	2.527
FUR_SZ_TIME	-0.1979	0.199	-0.994	0.320	-0.588	0.192
ROLLING_TEMP_T5	2.6350	0.476	5.532	0.000	1.701	3.569
ROLLING_DESCALING	-0.7007	0.245	-2.865	0.004	-1.180	-0.221

- R-squared: 42.54% (모델 설명력)

<참고> 표준화 상관 계수

<AxesSubplot:>



- 81.5%의 정확도

## (2) 의사결정나무

{'criterion': 'entropy', 'max\_depth': 5, 'min\_samples\_leaf': 10}

- GridSearchCV

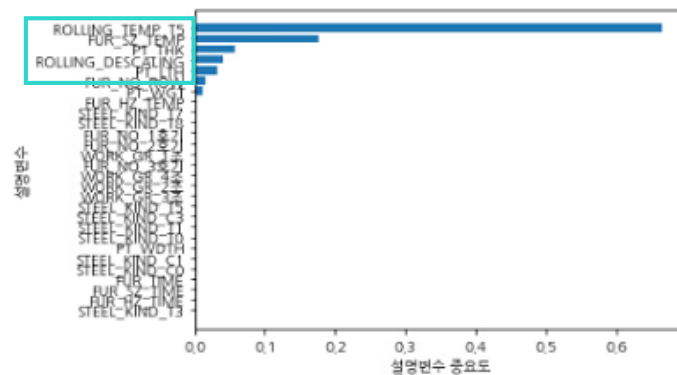
TrainAccuracy TestAccuracy

0 0.938492 0.944444

- 94.4%의 정확도

	precision	recall	f1-score	support
0	0.924	1.000	0.960	145
1	1.000	0.831	0.908	71
accuracy			0.944	216
macro avg	0.962	0.915	0.934	216
weighted avg	0.949	0.944	0.943	216

- 설명변수 및 중요도



## (3) 랜덤포레스트

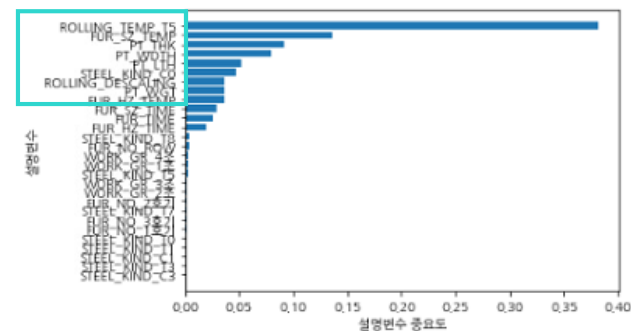
Accuracy on training set : 1.000

Accuracy on test set : 0.935

- Train은 100%, Test는 93.5%

	precision	recall	f1-score	support
0	0.924	1.000	0.960	145
1	1.000	0.831	0.908	71
accuracy			0.944	216
macro avg	0.962	0.915	0.934	216
weighted avg	0.949	0.944	0.943	216

- 설명변수 및 중요도



# 모델링

## 결과적으로 그래디언트부스팅 모델 성능이 가장 우수

주요 변수: 압연온도 / 가열로 균열대 시간 / 압연 중 Descaling 횟수 / Plate 두께

\* 앞선 탐색적 분석에서 가열로 '균열대 온도'와 '압연온도'가 완벽한 정비례 관계로, '균열대 온도' 또한 주요 인자라 볼 수 있다.

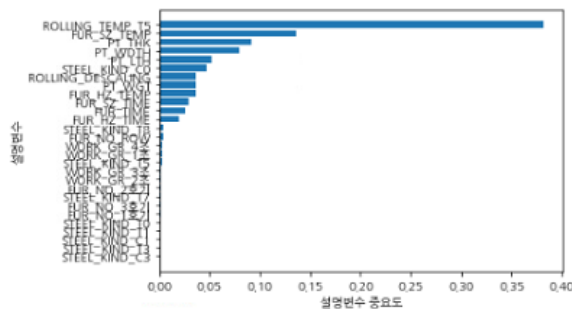
### (4) 그래디언트부스팅

Accuracy on training set : 0.986  
Accuracy on test set : 0.944

- Train은 98.6%, Test는 94.4%

	precision	recall	f1-score	support
0	0.924	1.000	0.960	145
1	1.000	0.831	0.908	71
accuracy			0.944	216
macro avg	0.962	0.915	0.934	216
weighted avg	0.949	0.944	0.943	216

- 설명변수 및 중요도



### (5) SVM

Accuracy on training set: 0.683  
Accuracy on test set: 0.671

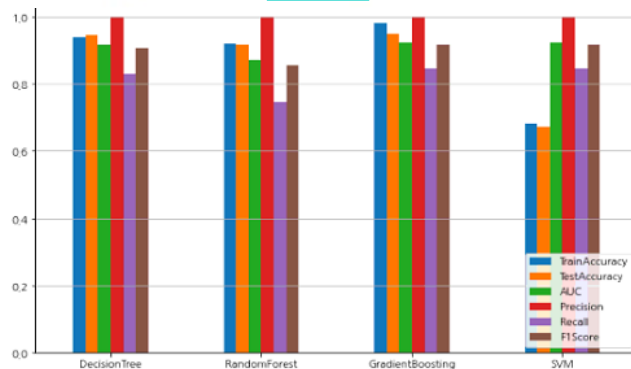
- Train은 68.3%, Test는 67.1%

Accuracy on training set: 0.875  
Accuracy on test set: 0.829

- Train은 87.5%, Test는 82.9%

### 모델평가

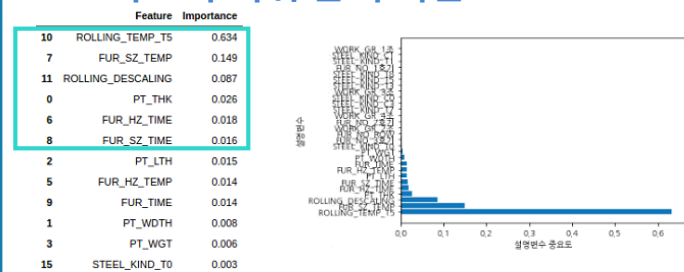
	TrainAccuracy	TestAccuracy	AUC	Precision	Recall	F1Score
DecisionTree	0.938	0.944	0.915	1.0	0.831	0.908
RandomForest	0.919	0.917	0.873	1.0	0.746	0.855
GradientBoosting	0.982	0.949	0.923	1.0	0.845	0.916
SVM	0.683	0.671	0.923	1.0	0.845	0.916



- Gradient Boosting이 가장 우수  
: 94.9%의 정확도를 보인다.

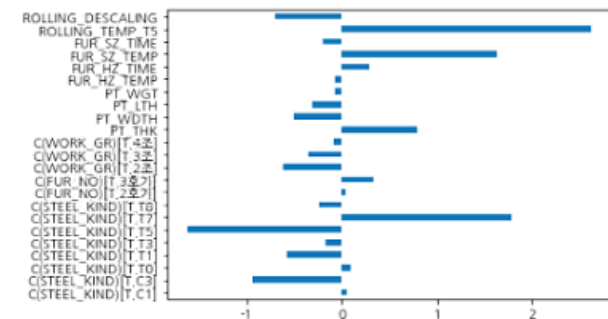
### 주요 변수

- 로지스틱 회귀 분석 기준



- Train은 87.5%, Test는 82.9%

- 그래디언트부스팅 기준



- Train은 87.5%, Test는 82.9%



# 결론

## 결론

- 모델링

- 로지스틱회귀 / 의사결정나무 / 랜덤포레스트 /  
그래디언트 부스팅 / SVM 실시

- 그래디언트 부스팅이 가장 높은 정확도를 보임  
: 94.9% 이상의 모델 정확도로 실무에서도 사용 가능

- 그래디언트 부스팅의 주요 변수 (상위 6개)

	Feature	Importance
10	ROLLING_TEMP_T5	0.634
7	FUR_SZ_TEMP	0.149
11	ROLLING_DESCALING	0.087
0	PT_THK	0.026
6	FUR_HZ_TIME	0.018
8	FUR_SZ_TIME	0.016

## 제안 및 평가

- 압연공장 SCALE 발생 감소 제안

1. 가열대 온도 감소
2. 가열로 균열대 온도 감소
3. 압연 중 Descaling 횟수를 증가
4. Plate 두께가 얇은 제품 위주 생산

- (예상 가설)

탐색적 분석 시, Histogram을 통해 Plate 폭 / 가열로 가열대 온도/  
가열로 균열대 온도 / 추출 온도와 불량률의 상관성을 확인  
→ 그래디언트 부스팅 주요 변수와 어느 정도 일치

- (기대 효과)

해당 생성 모델로 향후 불량률 예측을 통해  
불량 제품 사전 방지 가능

An aerial photograph of a large industrial complex, possibly a refinery or chemical plant, captured during the "blue hour" of twilight. The sky is filled with dramatic, dark clouds tinged with orange and pink from the setting or rising sun. The facility itself is a dense network of structures, including tall distillation columns, large storage tanks, and a complex system of pipes and walkways. Numerous artificial lights are illuminated throughout the plant, creating a warm, golden glow that contrasts with the cool tones of the evening sky. In the background, a body of water is visible, with some distant lights and structures on the horizon. The overall scene conveys a sense of industrial activity and scale.

**Thank You**