

후판공정 Scale 불량 원인 분석 보고서

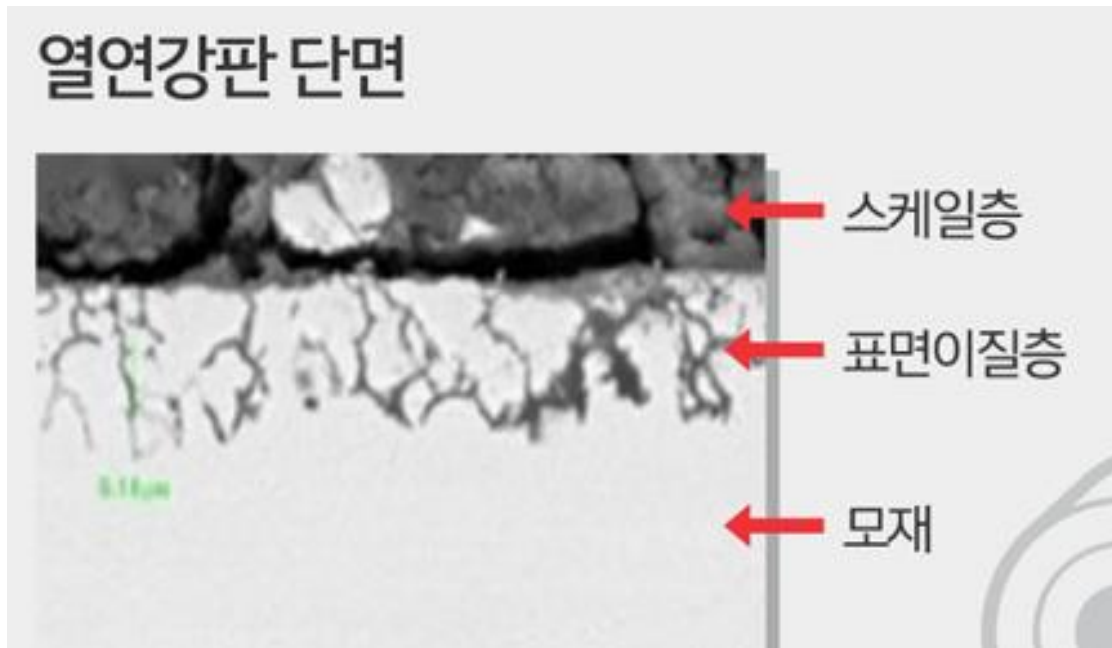
C반 2조 김병우/김준혁/김예원/배빛나리/신지우/심정욱

목차

CONTETNS

- 01 과제 정의
- 02 분석 계획
- 03 데이터 현황
- 04 탐색적 분석
- 05 모델링 요약
- 06 결론
- 07 느낀 점

1. 스케일이란 무엇인가?



- 강판의 스케일(scale)은 강철 표면에 형성된 산화철 층
- 산화층은 고온에서 강철이 공기 중의 산소와 반응할 때 생기며, 강철 제조 과정에서 가열 및 냉각 과정 중에 형성된다.
- 스케일은 강판의 표면 품질과 외관에 영향을 줄 수 있다.

➔ 강판의 스케일을 제거하는 공정이 필요.

과제 정의

1. POS Scale 불량 데이터에 기술통계 적용 계획 수립
2. Scale 불량 데이터의 속성을 특정하여 통계량을 사용해 정리, 요약, 설명
3. 정보화 시킨 데이터를 이용하여 어떤 핵심영향인자가 Scale 불량에 많은 영향을 끼치는지 파악
4. Scale 불량에 영향을 끼치는 인자들을 기준으로 Scale 불량을 예방하고 현장 개선안 도출
5. 결과에 이르기까지의 순서, 과정을 기술하고 시행착오를 통해 배운 내용을 기술.

목표

후판 제품 Scale 불량 원인을 찾고 목표 불량 발생률 0% 유지를 위하여

개선된 프로세스와 예방조치를 통해 핵심영향인자 도출 모델개선 프로세스 개발

1. 도메인 정보 파악



〈데이터 열 목록〉

#	Column
0	plate_no
1	rolling_date
2	scale
3	spec_long
4	spec_country
5	steel_kind
6	pt_thick
7	pt_width
8	pt_length
9	hsb
10	fur_no
11	fur_input_row
12	fur_heat_temp
13	fur_heat_time
14	fur_soak_temp
15	fur_soak_time
16	fur_total_time
17	fur_ex_temp
18	rolling_method
19	rolling_temp
20	descaling_count
21	work_group

[데이터 종류 설명]

• 목표 변수

Scale : 산화철 불량 (목표변수)

• 설명 변수 중 일부

Plate_no : 제품의 번호 (Id)

Rolling_date : 작업을 수행한 시각 (날짜형)

Spec_long : 제품 규격

Spec_country : 제품 규격 기준국

Steel_kind : 강종

Pt_thick : 후판의 지시 두께

Pt_width : 후판의 지시 폭

Pt_Length : 후판의 지시 길이

HSB : HSB (Hot Scale Braker; 스케일 제거 공정) 적용 여부

Fur_no : 가열로 호기 (몇번 가열로에서 작업을 수행했는가)

Fur_input_row : 가열로의 장입 열

Fur_heat_temp : 가열로 가열대 소재 온도

Fur_heat_time : 가열로의 가열대 재로 시간

Fur_soak_temp : 가열로 가열대 재로 시간

Rolling_method : 압연방법 (TMCP/CR)

* 이하 생략

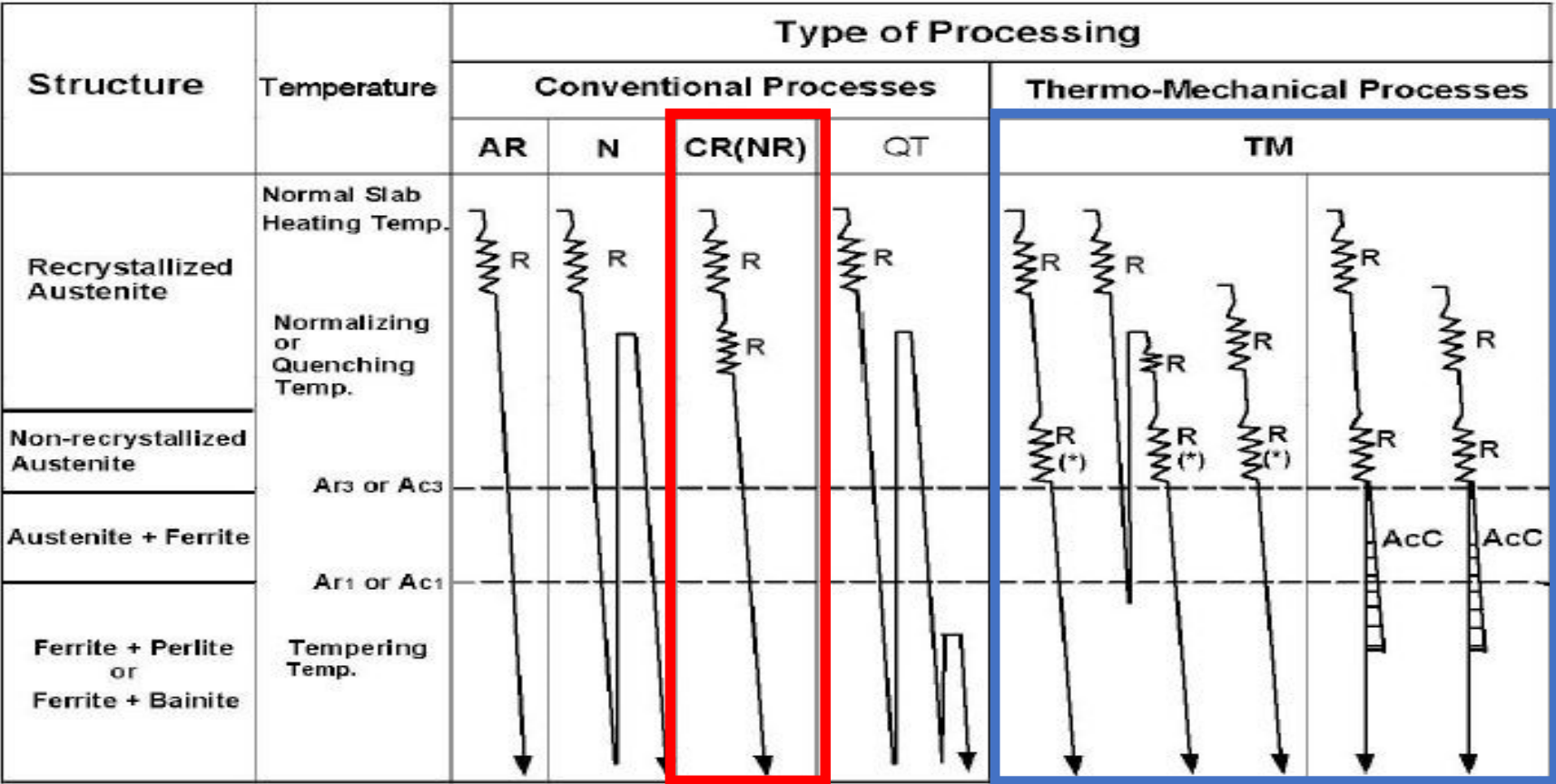
2. 변수 상세 설명



[steel_kind]

강종	
T(티타늄)	C(탄소강)
티타늄은 고온에서의 산화 반응으로 인해 형성 되는 Scale이 밀도가 높고 견고한 보호층을 형성한다.	탄소강에서 형성되는 스케일은 상대적으로 더 두껍고 불규칙한 구조를 가진다.

[rolling_method]



CR(Controlled Rolling) : 강철의 기계적 성질을 얻기 위해 냉각 속도를 조절하는 압연 방법이다. TMCP보다 냉각속도가 상대적으로 느리다.

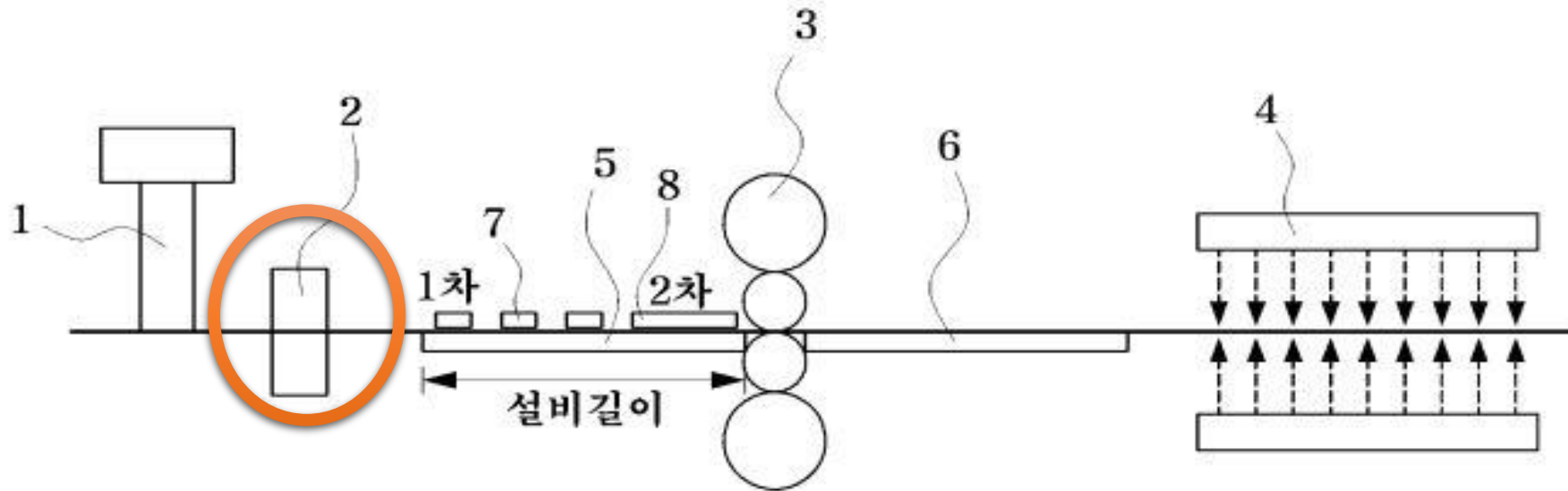
TMCP(Thermo-Mechanical Control Process) : 압연 후 강철을 빠르게 냉각시켜서 마르텐사이트 변태를 억제하고 강도를 향상시킨다.

2. 변수 상세 설명

[HSB, hot scale breaker]

열간 압연 공정 초기에 금속 표면에 형성된 스케일(산화층)을 제거하는 데 사용한다.

스케일을 사전에 제거해서 롤과 금속 사이의 마찰을 줄이고, 압연 공정의 효율을 향상시킨다.

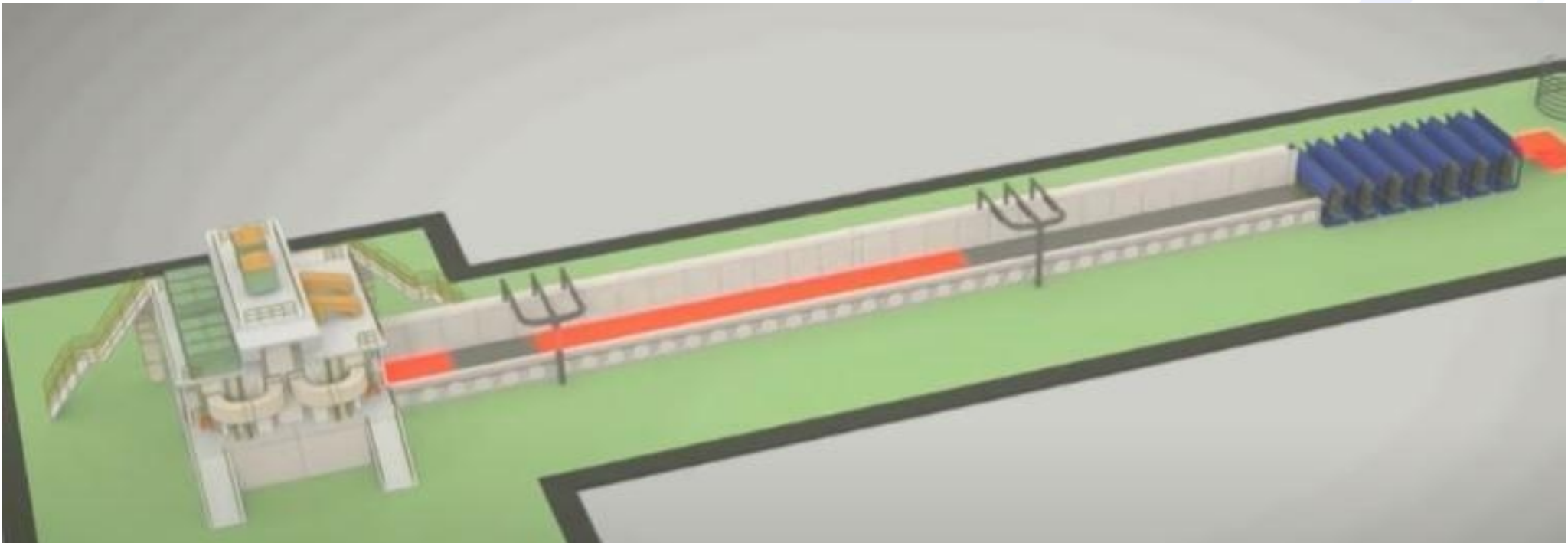


1. 가열로 2. HSB 3. 압연기 4. 가속냉각기 5,6 롤러 테이블

2. 변수 상세 설명

[fur_ex_temp]

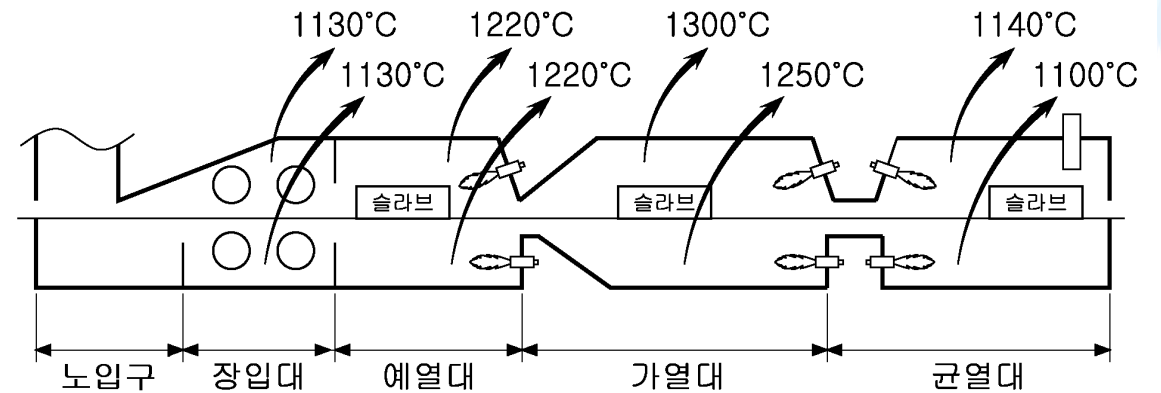
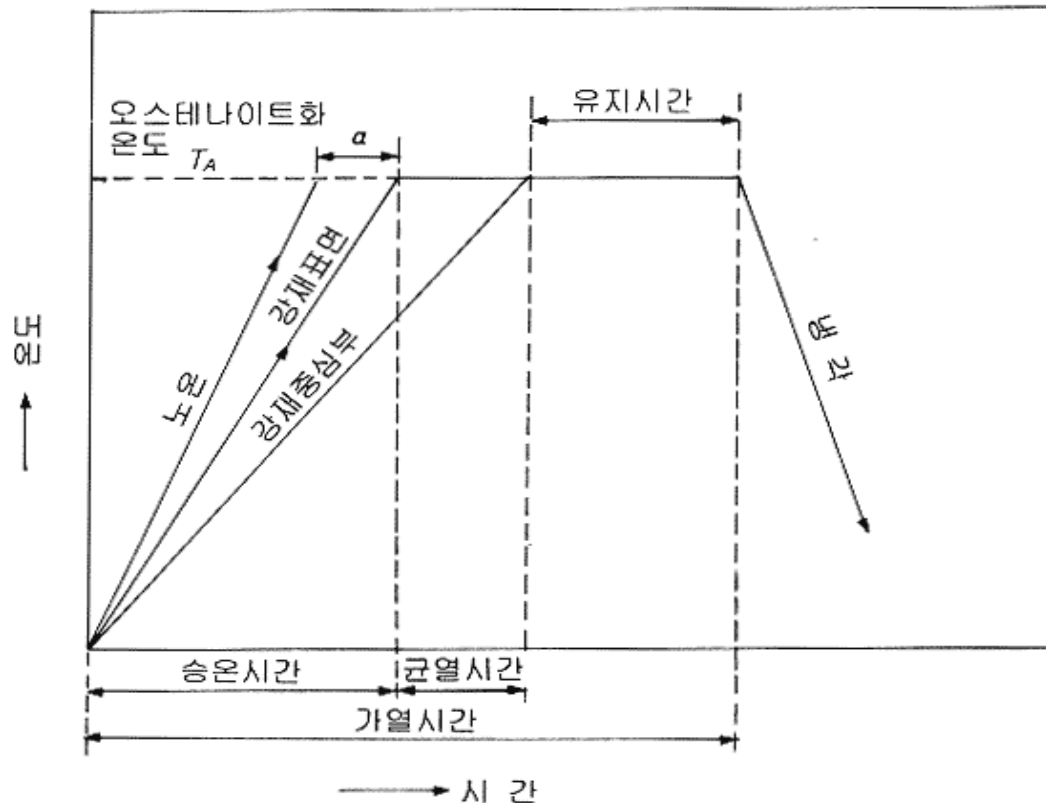
가열로에서 후판을 추출하는 온도가 높으면, 금속의 산화 반응이 빨라져 더 두꺼운 scale이 형성된다.
반대로 추출 온도가 낮으면, 상대적으로 얇은 scale이 형성된다.



2. 변수 상세 설명

[fur_total_time]

총 재로시간 = 예열시간(승온시간) + 균열시간 + 대기시간



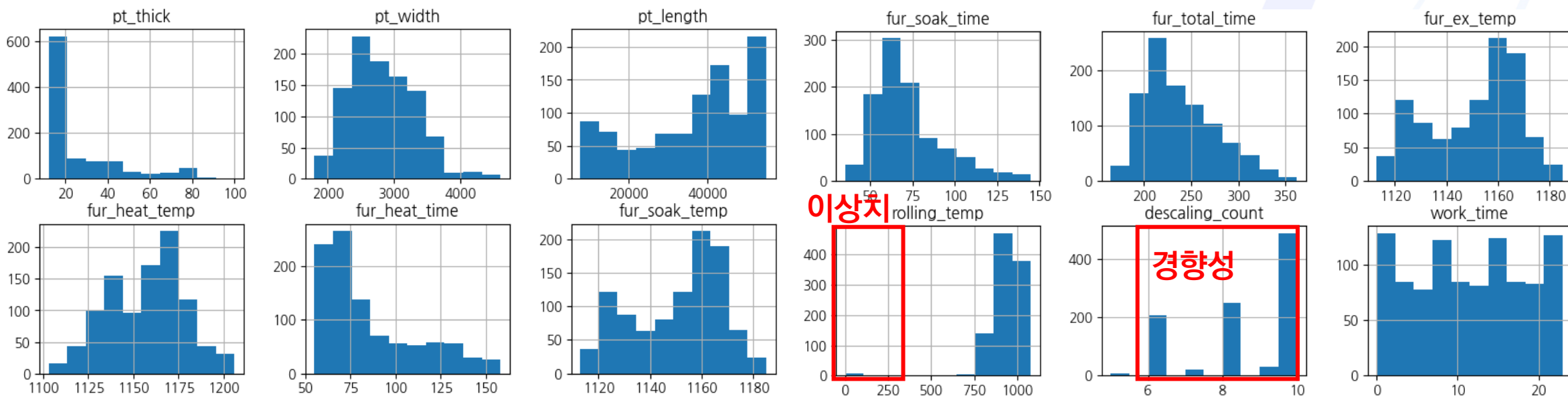
출처 : 논문 - A p i 재 열함량 일정화 조업방법

- 예열대 : 소재를 천천히 가열하여 내부와 외부의 온도 차이를 최소화함으로써 내부 응력을 줄인다.
- 가열대 : 압연에 필요한 최적 온도까지 가열한다.
- 균열대 : 가열된 소재를 압연 전에 온도를 균일하게 분포시킨다.

1. 요약통계량

〈변수별 요약통계량 및 히스토그램을 통한 데이터 분포 파악〉

	pt_thick	pt_width	pt_length	fur_heat_temp	fur_heat_time	fur_soak_temp	fur_soak_time	fur_total_time	fur_ex_temp	rolling_temp	descaling_count	work_time
count	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000	1000.00000
mean	26.78200	2831.900000	36788.200000	1157.245000	85.972000	1150.928000	71.720000	238.589000	1150.928000	934.637000	8.557000	11.493000
std	18.13757	494.081478	13912.387116	21.245007	26.346297	17.344384	20.602137	38.194828	17.344384	96.598015	1.604158	6.970979
min	12.00000	1800.000000	7900.000000	1103.000000	55.000000	1113.000000	35.000000	165.000000	1113.000000	0.000000	5.000000	0.000000
25%	15.00000	2500.000000	26650.000000	1140.000000	66.000000	1135.750000	57.750000	210.000000	1135.750000	893.750000	8.000000	5.000000
50%	19.00000	2800.000000	40400.000000	1159.000000	75.000000	1156.000000	66.000000	230.000000	1156.000000	948.000000	9.000000	11.500000
75%	34.00000	3100.000000	49100.000000	1173.000000	102.250000	1164.000000	81.000000	263.000000	1164.000000	991.000000	10.000000	17.250000
max	100.00000	4600.000000	54900.000000	1206.000000	158.000000	1185.000000	145.000000	362.000000	1185.000000	1078.000000	10.000000	23.000000

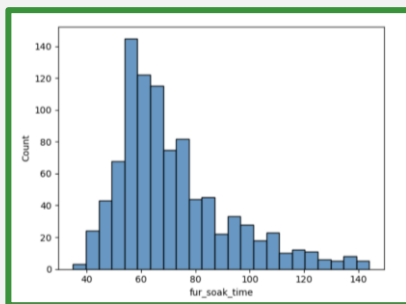
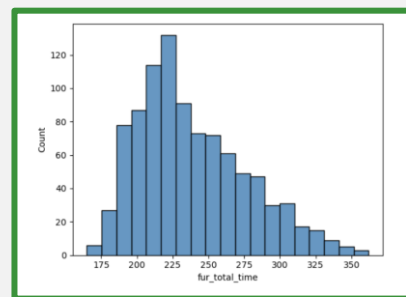


➔ 요약통계량을 통해 목표변수와 설명변수의 특징을, 히스토그램을 통해 이상치 존재 여부와 개형을 파악했다.

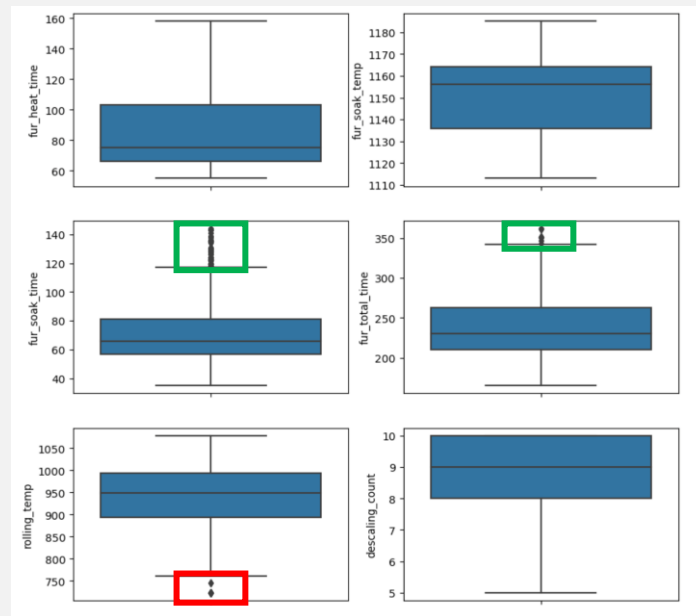
2. 이상치 분석 및 제거

〈설명 변수들의 Box Plot과 Histogram〉

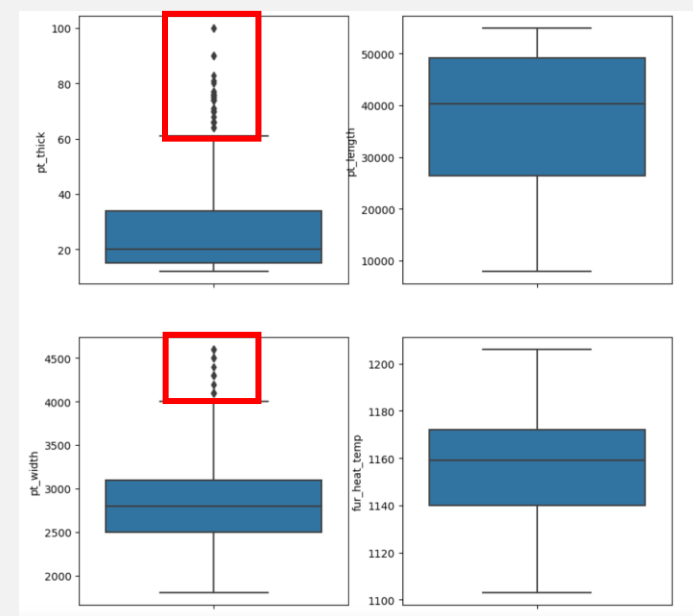
설명변수들에 대해 Box Plot 과 Histogram 을 통해 이상치가 존재하는지 검증해본다.



균열대 재료 시간과 총 재료시간의 분포를 확인해본 결과, 정규분포에 가까워 이상치를 제거하지 않기로 했다.



압연온도가 0℃ 일리는 없기에 이상치로 판단하여 제거



후판의 지시폭, 두께변수에서 Q3를 넘는 이상치들이 발견되었지만, 해당 값들은 발주 측에서 평균보다 큰 규격을 요구한 것이므로 이상치라고 판단할 수 없다.

2. 이상치 분석 및 제거

〈분석에 영향이 적은 연속형, 범주형 변수 제거〉

분석에 포함되지 않는 큰 의미 없는 변수 제거

Plate_no
Plate 번호Spec_long
제품 규격Fur_ex_temp
가열로 추출온도

plate_no
PLT_1001
PLT_1002
PLT_1003
PLT_1004

: Id

spec_long
AB/EH32-TM
AB/EH32-TM
NV-E36-TM
NV-E36-TM

: 국가별 규격+기법

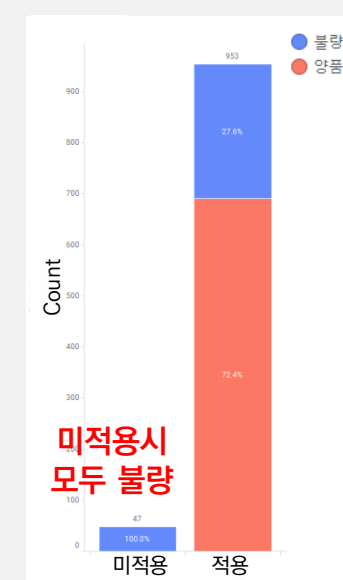
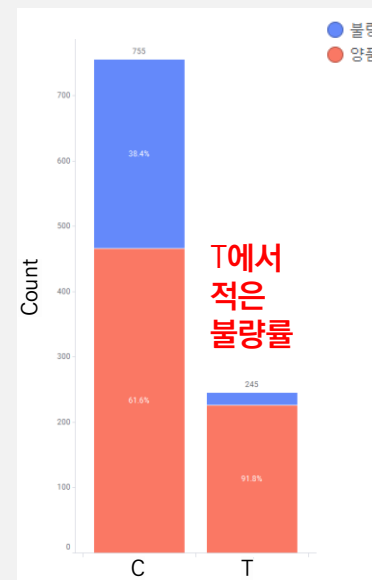
fur_soak_temp	fur_ex_temp
1133	1133
1135	1135
1121	1121
1127	1127

: 동일한 변수

PLT_1995
PLT_1996
PLT_1997
PLT_1998
PLT_1999
PLT_2000

LR-A
BV-A
LR-A
AB/AH32
NV-A32
GL-A32

층별화를 위한 변수 제거

Steel_kind
강종 (C/T)HSB
적용여부 (0/1)

Plate_no, Spec_long, Fur_ex_temp 열은 drop한다. 티타늄(T)은 제거하고 탄소강(C)만 분석, HSB 미적용은 제거, HSB 적용만 분석을 진행한다.

3. 데이터 특성 파악

〈시계열분석〉

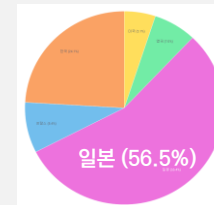
[작업조별 Rolling_Date에 따른 불량률 그래프]



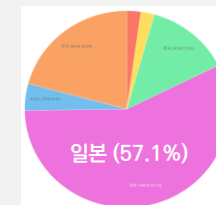
[Rolling_Date에 따른 불량률 그래프]



[불량 연속 발생 일자의 주문 국가]



[1조/1월 4일]



[2조/1월 4일]

불량률이 가장 높았던 날에
일본 발주 제품이 가장 큰
비율로 생산되었다.

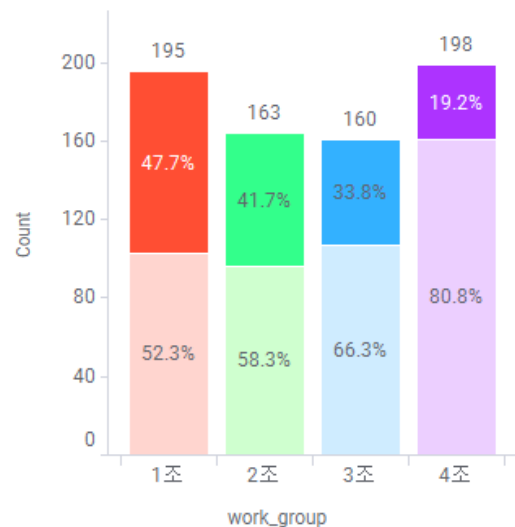
➔ 특정 일자 (2023-01-04)에 연속 불량 발생했고, 시계열로 불량 발생 현황을 확인해본 결과도 동일했다. 이는 해당 시기에 설비 고장이나 예측불가한 이상 원인이 있었음을 시사한다. 해당 일자의 발주 국가 분석 결과 일본에서 발주한 제품을 생산하면서 불량이 많이 발생했다는 것을 알 수 있다.

3. 데이터 특성 파악

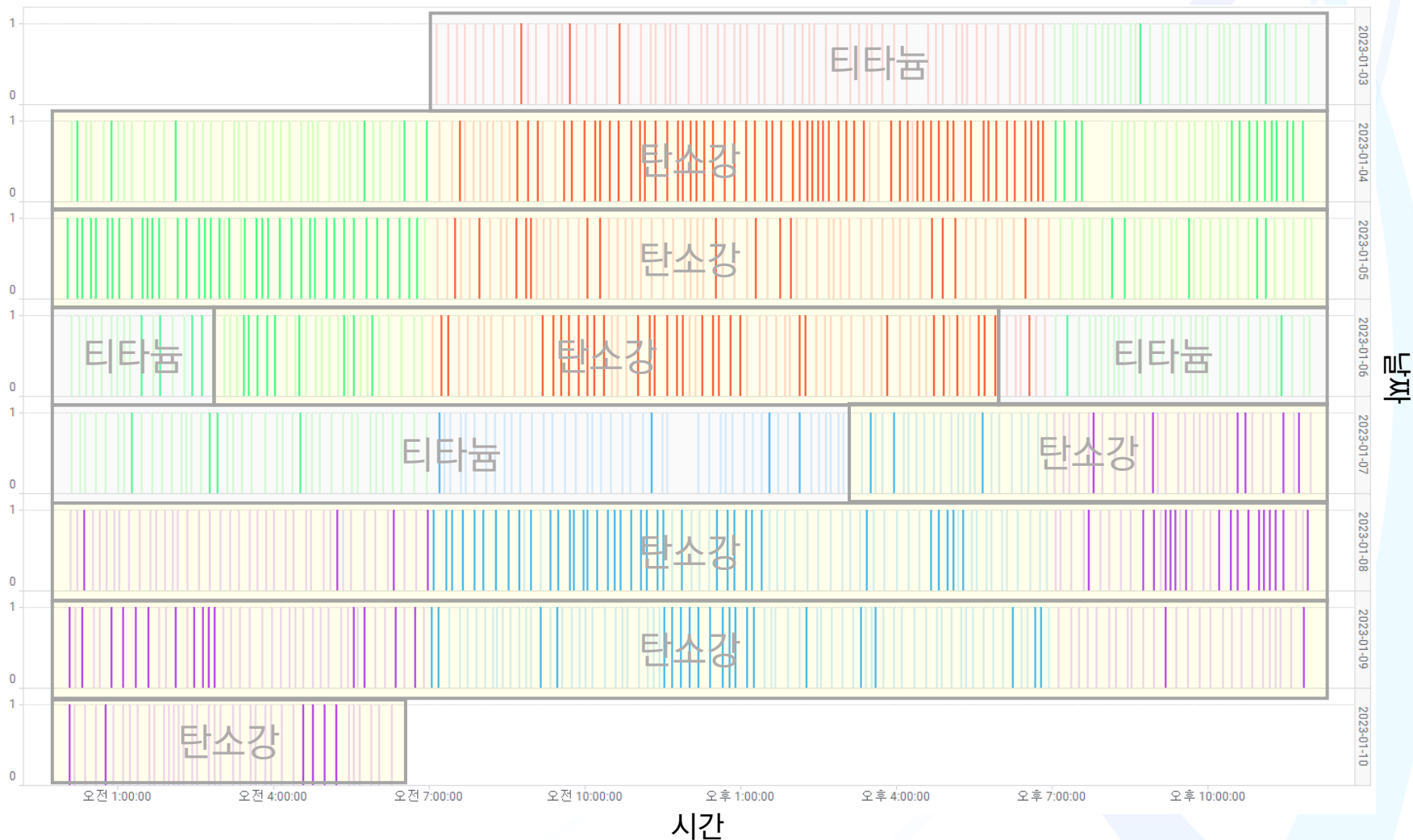
〈시계열분석을 통한 작업조별 시각화〉

조별
불량
표시

- 1조 » 불량
- 1조 » 양품
- 2조 » 불량
- 2조 » 양품
- 3조 » 불량
- 3조 » 양품
- 4조 » 불량
- 4조 » 양품



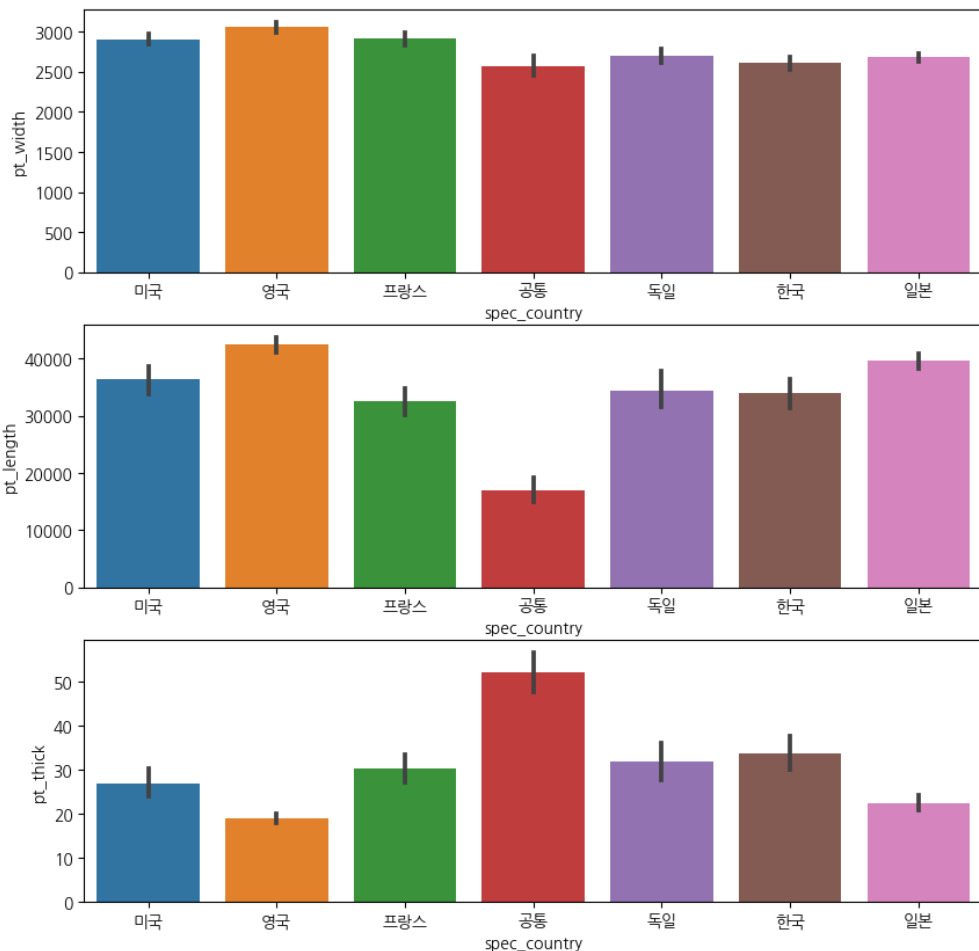
아래는 티타늄까지 고려한 날짜와 시간에 따른 조별 불량률 표이다. 연한 색은 양품, 진한 색은 불량을 나타낸다.



3. 데이터 특성 파악

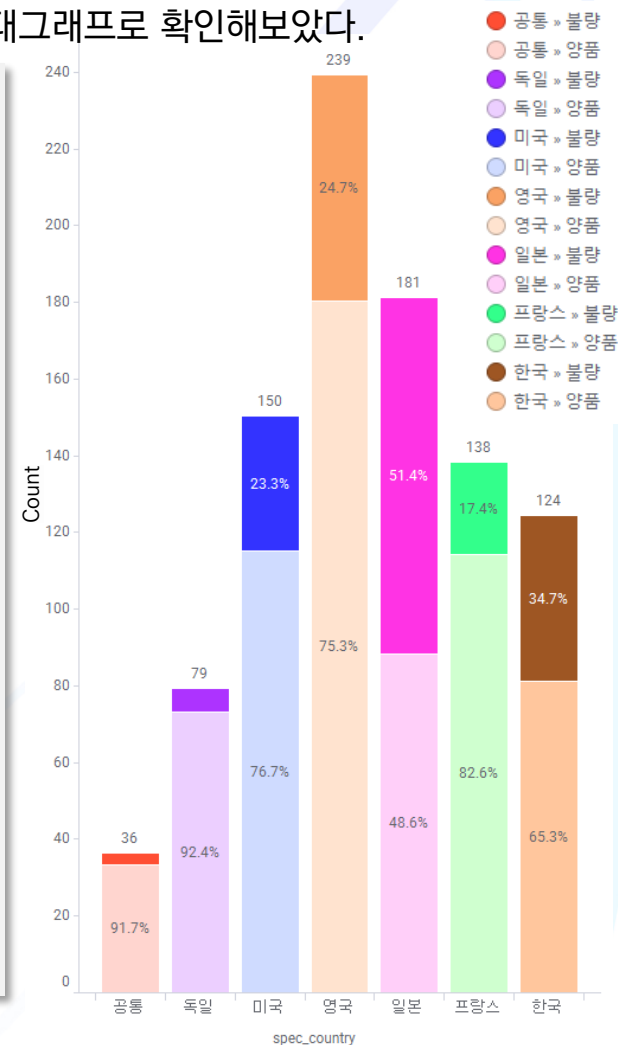
〈국가별 분석〉

국가별 발주 스펙에 차이가 있다고 가설을 세우고, 국가별 지시 규격과 총 주문 대비 양품과 불량품 발생 여부를 막대그래프로 확인해보았다.



좌측의 막대그래프로부터 국가별 주문 스펙(지시폭, 지시길이, 지시두께)을 확인해 본 결과, 지시폭(pt_width)에서는 국가별로 명확한 차이를 확인하기 어려웠지만, 지시 길이(pt_length)에 있어서는 **영국, 일본, 미국** 분류에서 큰 값과, **공통** 분류에서 작은 값을 확인할 수 있었다. 또한 지시두께(pt_thick)에 있어서는 **영국, 일본, 미국**에서 작은 값과, **공통** 분류에서 큰 값을 확인할 수 있었다. 따라서, 공통분류에 비해서 **영국, 일본, 미국**은 생산이 까다로운 주문이 주를 이뤘다고 할 수 있다.

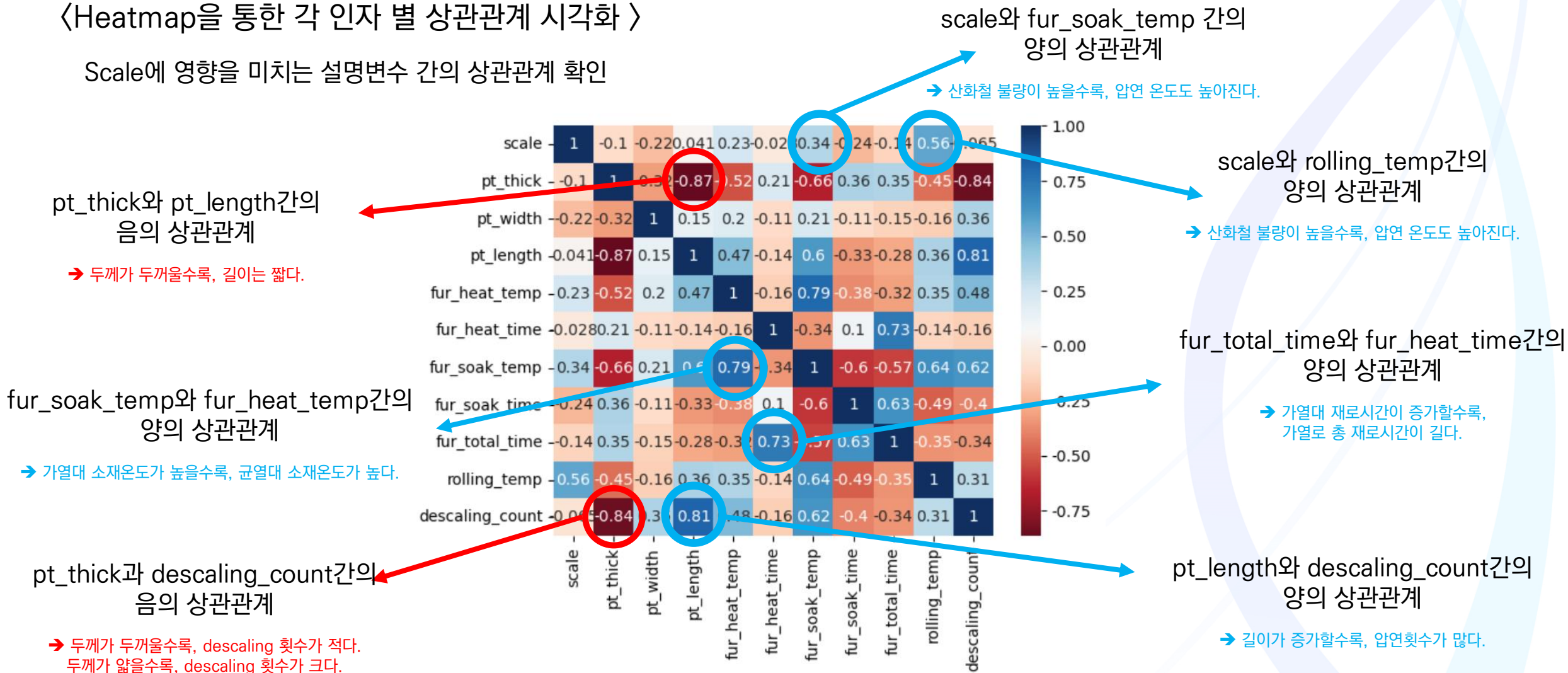
우측 불량률 막대그래프에서도 불량률이 **일본, 한국, 영국, 미국**에서 높으며, **공통** 분류에서 낮은 것을 확인할 수 있다.



4. 상관관계 시각화

〈Heatmap을 통한 각 인자 별 상관관계 시각화〉

Scale에 영향을 미치는 설명변수 간의 상관관계 확인



1. Spearman Correlation 검정

〈가열로 호기, 가열로 장입열, 작업조가 Scale에 영향을 미치는지 여부〉

[work_group 과 scale]

Corr = -0.126로 상관성이 없다

작업조와 불량률은 관계 없다.

[fur_no 와 scale]

Corr = 0.035 로 상관성이 없다.

가열로 호기와 불량률은 관계 없다.

[fur_input_row 와 scale]

Corr = 0.038 로 상관성이 없다

가열로의 장입열과 불량률은 관계 없다.

두 열 간의 스피어만 상관관계 계산

```
for i in ["work_group", "fur_no", "fur_input_row"] :
    correlation, p_value = spearmanr(df['y'], df[i])

    # 상관관계 출력
    print("Spearman Correlation:", correlation)
    print("p-value:", p_value)

    # 상관관계 해석
    print(i)
    if abs(correlation) < 0.5:
        print("Variables are not significantly correlated.")
    else:
        print("Variables are significantly correlated.")
    print()
```

Spearman coefficient는 범주형 변수 간 상관성을 보는 지표이다.
일반적으로 절대값 0.5이하의 계수는 상관성이 없음을 시사한다.

```
Spearman Correlation: -0.1261895042252613
p-value: 9.871798017002286e-05
work_group
Variables are not significantly correlated.
```

```
Spearman Correlation: 0.03512270977910579
p-value: 0.28025370844481884
fur_no
Variables are not significantly correlated.
```

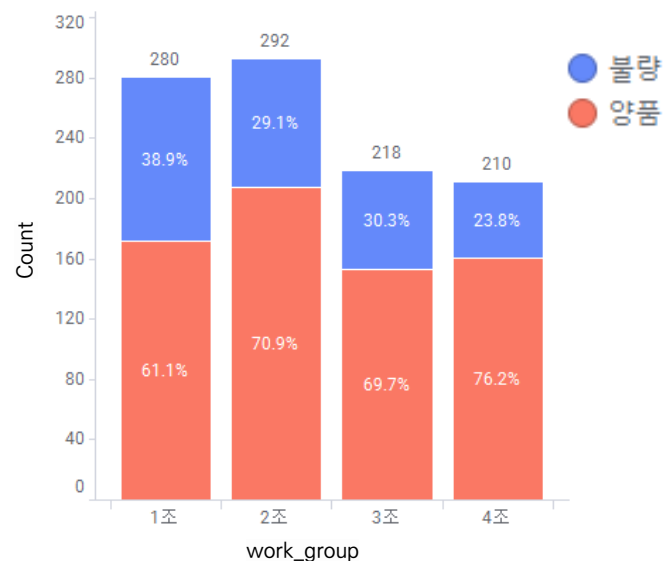
```
Spearman Correlation: -0.038387142538428155
p-value: 0.23793009070477011
fur_input_row
Variables are not significantly correlated.
```

2. Bar Plot

〈가열로 호기, 가열로 장입열, 작업조가 Scale에 영향을 미치는지 여부〉

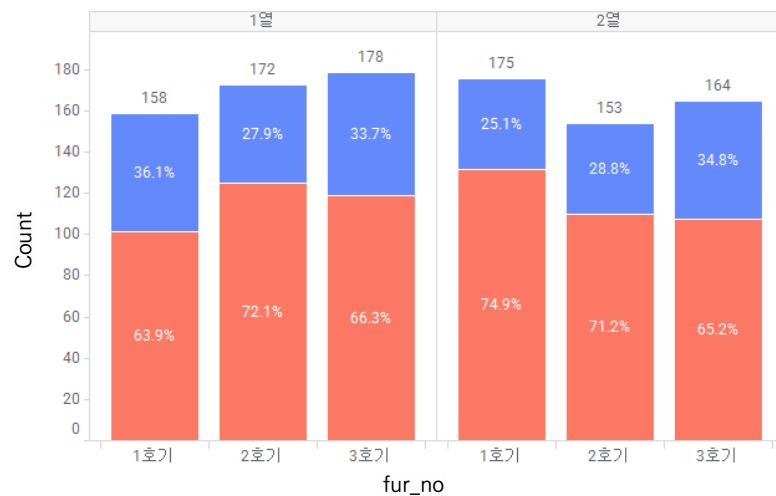
[work_group 과 scale]

작업조와 불량률은 관계 없다.



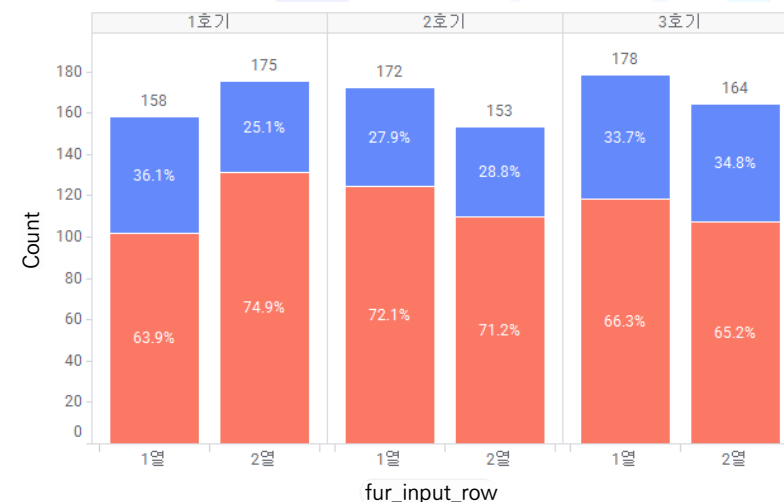
[fur_no 와 scale]

가열로 호기와 불량률은 관계 없다.



[fur_input_row 와 scale]

가열로의 장입열과 불량률은 관계 없다.



그래프를 통해 확인해보아도, 작업조, 가열로 호기, 가열로의 장입열에 따른 불량률에 있어, 눈에 띄게 큰 차이는 없다는 것을 알 수 있다.

1. t-Test 가설 검정

〈불량률에 영향을 미치는 인자 여부〉

[pt_length와 scale]

P-value < 0.05 이므로 귀무가설을 따른다.**지시길이**가 불량률에 영향을 미친다.

```
from scipy.stats import ttest_ind

# 'scale' 열을 기준으로 양품과 불량품 데이터를 필터링
good_quality = df.query('y == 0')['pt_length']
bad_quality = df.query('y == 1')['pt_length']

# 독립 표본 t-검정 수행
t_stat, p_value = ttest_ind(good_quality, bad_quality)

# 결과 출력
print("t_stat: ", t_stat, "\np_value: ", p_value)
```

귀무 가설(H_0): 양품과 불량품 그룹의 `pt_length` 평균은 같다.• $H_0: \mu_{\text{양품}} = \mu_{\text{불량품}}$ 대립 가설(H_a): 양품과 불량품 그룹의 `pt_length` 평균은 다르다.• $H_a: \mu_{\text{양품}} \neq \mu_{\text{불량품}}$

```
t_stat: 11.338291670123235
p_value: 1.6263936225266937e-27
```

[pt_width와 scale]

P-value < 0.05 이므로 귀무가설을 따른다.**지시폭**이 불량률에 영향을 미친다.

```
from scipy.stats import ttest_ind

# 'scale' 열을 기준으로 양품과 불량품 데이터를 필터링
good_quality = df.query('y == 0')['pt_width']
bad_quality = df.query('y == 1')['pt_width']

# 독립 표본 t-검정 수행
t_stat, p_value = ttest_ind(good_quality, bad_quality)

# 결과 출력
print("t_stat: ", t_stat, "\np_value: ", p_value)
```

귀무 가설(H_0): 양품과 불량품 그룹의 `pt_width` 평균은 같다.• $H_0: \mu_{\text{양품}} = \mu_{\text{불량품}}$ 대립 가설(H_a): 양품과 불량품 그룹의 `pt_width` 평균은 다르다.• $H_a: \mu_{\text{양품}} \neq \mu_{\text{불량품}}$

```
t_stat: 10.510214186528485
p_value: 4.021886463549695e-24
```

[pt_thick와 scale]

P-value < 0.05 이므로 귀무가설을 따른다.**지시두께**가 불량률에 영향을 미친다.

```
from scipy.stats import ttest_ind

# 'scale' 열을 기준으로 양품과 불량품 데이터를 필터링
good_quality = df.query('y == 0')['pt_thick']
bad_quality = df.query('y == 1')['pt_thick']

# 독립 표본 t-검정 수행
t_stat, p_value = ttest_ind(good_quality, bad_quality)

# 결과 출력
print("t_stat: ", t_stat, "\np_value: ", p_value)
```

귀무 가설(H_0): 양품과 불량품 그룹의 `pt_thick` 평균은 같다.• $H_0: \mu_{\text{양품}} = \mu_{\text{불량품}}$ 대립 가설(H_a): 양품과 불량품 그룹의 `pt_thick` 평균은 다르다.• $H_a: \mu_{\text{양품}} \neq \mu_{\text{불량품}}$

```
t_stat: -12.600002033866051
p_value: 5.1656638082670644e-33
```

2. 다중선형회귀

〈온도 저하량 공식 검증〉

금속의 온도 저하량은 아래의 공식을 따른다고 한다. 다중선형 회귀를 활용하여 회귀식이 공식을 따르는지 검증한다.

[온도 저하량 공식]

온도 저하량[℃] = 11.7 + 0.000427 × 소재의 폭[mm] - 0.0383 × 소재의 두께[mm] + 0.00547 × 조압연 종료시의 소재 표면온도[℃]

〈다중 선형 회귀 결과〉

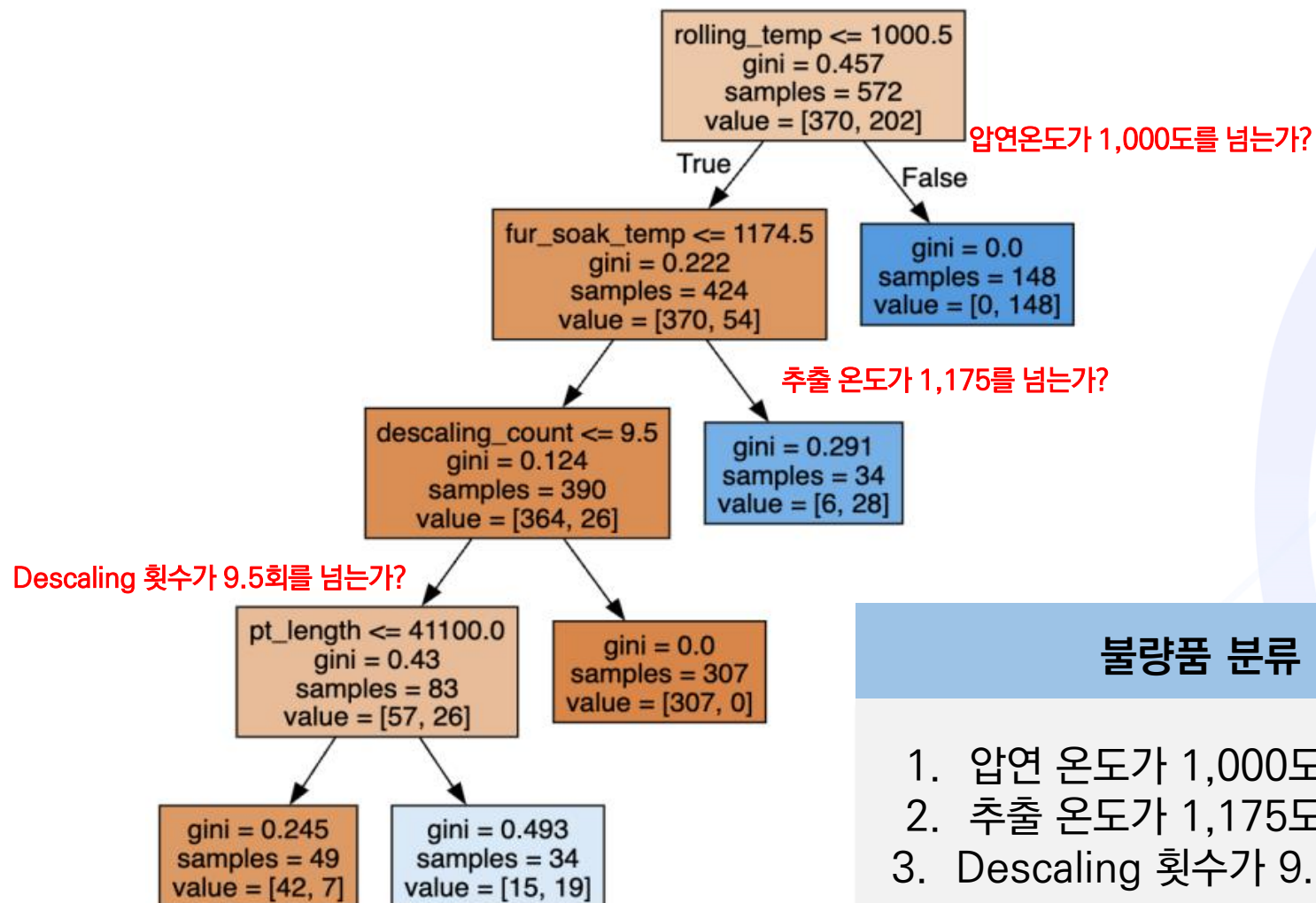
OLS Regression Results						
Dep. Variable:	temp_dif	R-squared:	0.008			
Model:	OLS	Adj. R-squared:	0.004			
Method:	Least Squares	F-statistic:	1.841			
Date:	Tue, 12 Mar 2024	Prob (F-statistic):	0.138			
Time:	05:14:13	Log-Likelihood:	-2809.4			
No. Observations:	700	AIC:	5627.			
Df Residuals:	696	BIC:	5645.			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975
Intercept	60.4177	45.601	1.325	0.186	-29.114	149.95
pt_width	0.0010	0.001	0.897	0.370	-0.001	0.00
pt_thick	0.0255	0.039	0.650	0.516	-0.051	0.10
fur_ex_temp	-0.0498	0.039	-1.279	0.201	-0.126	0.02
Omnibus:	18.927	Durbin-Watson:	2.026			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	39.858			
Skew:	-0.047	Prob(JB):	2.21e-09			
Kurtosis:	4.165	Cond. No.	2.77e+05			

	Coef
Constant	60.417749
Pt_Width	0.000976
Pt_Thick	0.025503
Fur_ex_temp	-0.049769

- 소재의 폭(Pt_Width), 소재의 두께(Pt_Thick), 표면 온도(Fur_ex_temp), 온도저하량(Fur_ex_temp - Fur_soak_temp)을 활용하여서 다중선형회귀를 진행한 결과 차이가 다소 큰 공식을 얻었다.
- 설명력이 0.8%로 매우 낮고, 계수 또한 차이가 많이 나는 것을 고려하면 주어진 데이터로 검증하는 것은 어렵다는 결론을 얻을 수 있었다.

3. Decision Tree

〈트리 모델 시각화 및 해석 : Scale에 영향을 미치는 핵심인자〉



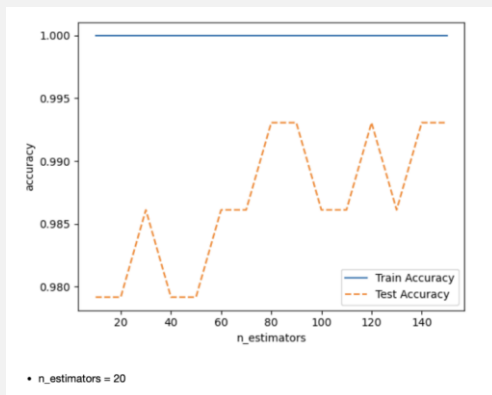
불량품 분류 프로세스

1. 압연 온도가 1,000도를 초과할 경우 불량
2. 추출 온도가 1,175도를 초과할 경우 불량
3. Descaling 횟수가 9.5회를 넘을 경우 양품

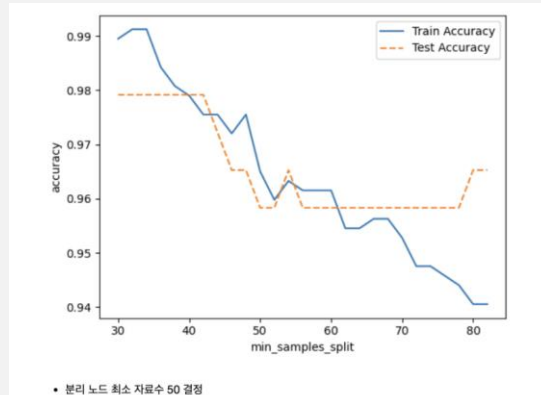
4. Random_Forest

〈Hyper Parameter Tuning〉

모델의 성능을 극대화하기 위해 Hyper Parameter를 변경해가면서 최적의 값을 찾는다.

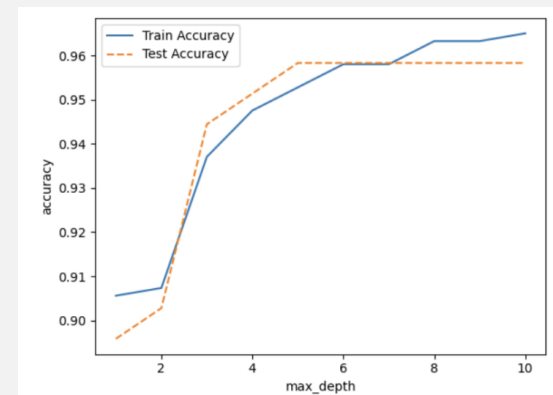


$n_estimators = 20$



$min_samples_split = 50$

분리 노드 수는 낮을수록 과적합되므로,
분리 노드 최소 자료수를 50으로 지정한다.



$max_depth = 2$

Depth가 2 이후부터는
과적합되므로 2로 지정한다.

최종 모델 선정

하이퍼파라미터 » $n_estimators = 20$, $min_samples_split = 50$, $max_depth = 2$

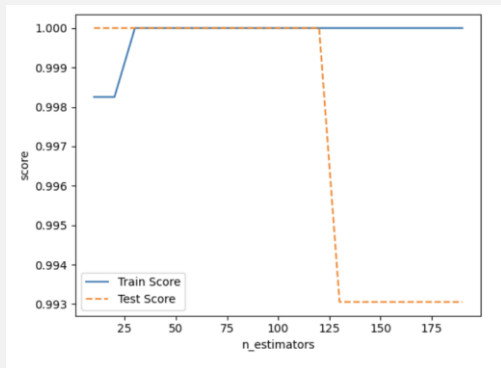
```
train data Accuray [0.9577039274924471]
test data Accuray [0.9543859649122807]
```

» 훈련 데이터와 테스트 데이터 모두에서 상당히 높은 정확도를 확인할 수 있었다.

5. XGBoost

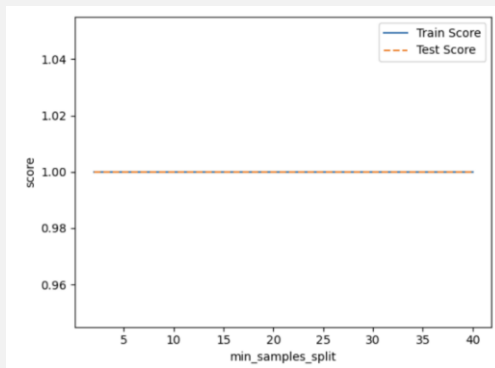
〈Hyper Parameter Tuning〉

모델의 성능을 극대화하기 위해 Hyper Parameter를 변경해가면서 최적의 값을 찾는다.



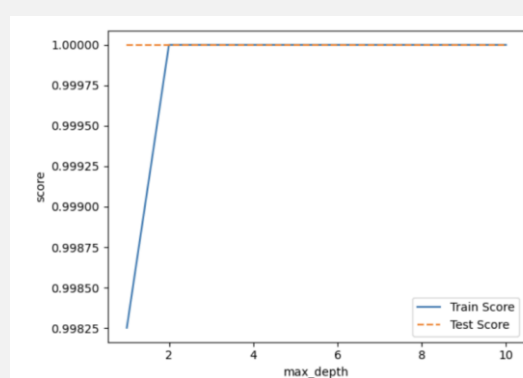
`n_estimators = 5`

처음에는 모델 성능에 영향을 주지 않는 것으로 보였지만, 좀 더 세부적으로 분석해 5로 지정한다.



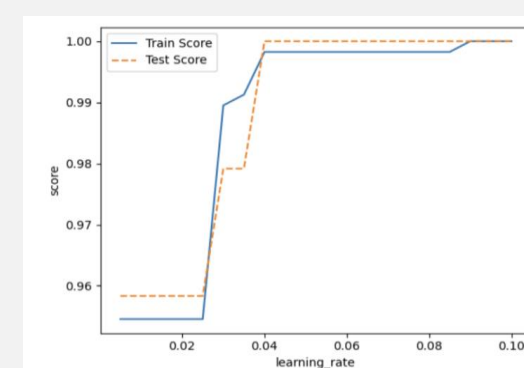
`min_samples_split` 미지정

앞사귀 노드 최소 자료수도 모델 성능에 영향을 주지 않기 때문에 지정하지 않는다.



`max_depth = 4`

Depth가 2일 때에는 과소적합하지만, 4 이후부터는 과적합되므로 4로 지정한다.



`Learning_rate = 0.075`

`Learning_rate`가 0.075 이후부터는 과적합되므로 0.075로 지정한다.

최종 모델 선정

하이퍼파라미터 » `n_estimators = 5`, `max_depth = 4`, `learning_rate = 0.075`

Score on train : 0.958
score on test : 0.958

» 훈련 데이터와 테스트 데이터 모두에서 상당히 높은 정확도를 확인할 수 있었다.

6. Logistic Regression

〈결과 확인〉

Optimization terminated successfully.
Current function value: 0.328851
Iterations 8

Logit Regression Results

```
=====
Dep. Variable:          y      No. Observations:          144
Model:                  Logit  Df Residuals:              137
Method:                  MLE   Df Model:              6
Date:                   Tue, 12 Mar 2024  Pseudo R-squ.:      0.4834
Time:                   19:44:41  Log-Likelihood:     -47.355
converged:               True   LL-Null:           -91.658
Covariance Type:         nonrobust LLR p-value:          5.898e-17
=====
```

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.8137	0.405	-4.482	0.000	-2.607	-1.021
pt_width	-0.7694	0.365	-2.106	0.035	-1.485	-0.053
pt_length	-0.4302	0.366	-1.176	0.240	-1.147	0.287
fur_heat_time	0.5227	0.319	1.637	0.102	-0.103	1.148
fur_soak_temp	1.5911	0.480	3.312	0.001	0.649	2.533
rolling_temp	1.3481	0.576	2.342	0.019	0.220	2.476
descaling_count	-0.6708	0.319	-2.106	0.035	-1.295	-0.046

로지스틱 회귀분석을 통해 p-value값을 기준으로 차례대로 변수를 지워가며 fur_soak_time, fur_total_time, spec_country 변수들이 유의미하지 않음을 확인했다.

Optimization terminated successfully.
Current function value: 0.328851
Iterations 8

Logit Regression Results

```
=====
Dep. Variable:          y      No. Observations:          144
Model:                  Logit  Df Residuals:              137
Method:                  MLE   Df Model:              6
Date:                   Tue, 12 Mar 2024  Pseudo R-squ.:      0.4834
Time:                   19:44:41  Log-Likelihood:     -47.355
converged:               True   LL-Null:           -91.658
Covariance Type:         nonrobust LLR p-value:          5.898e-17
=====
```

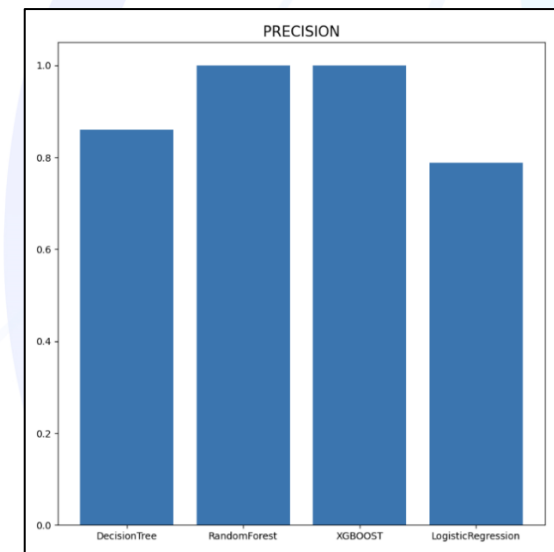
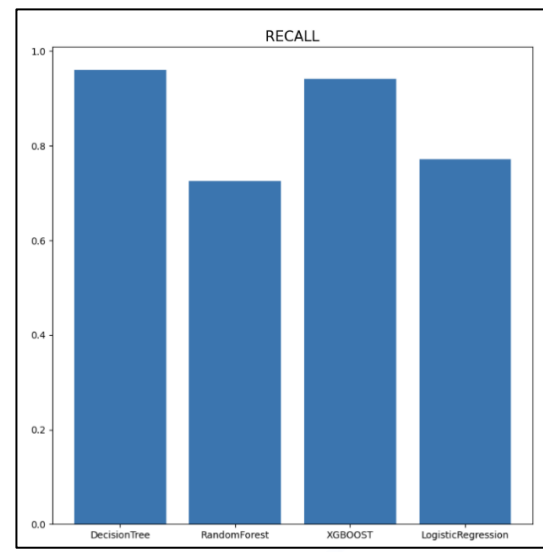
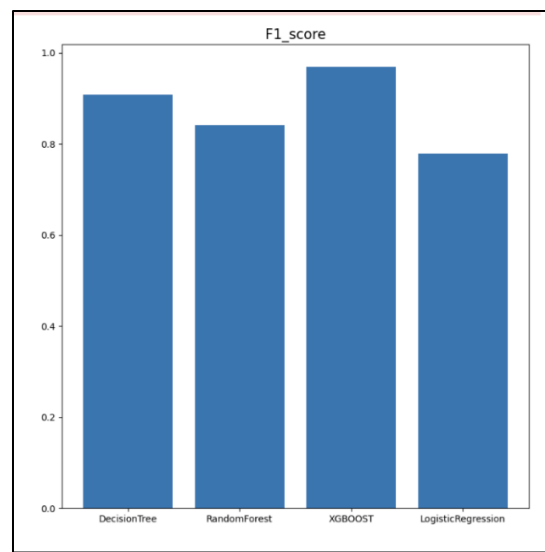
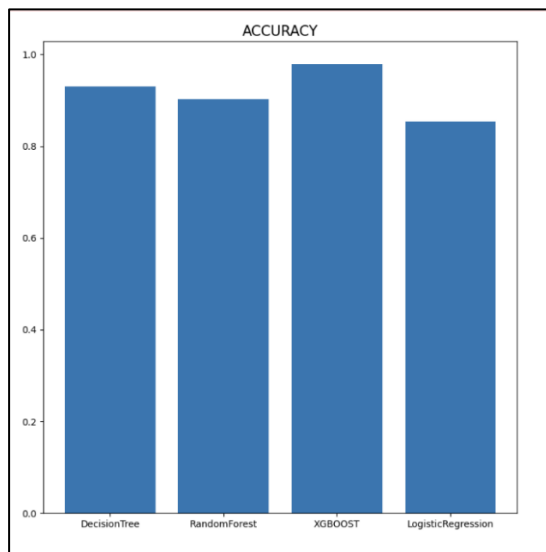
	coef	std err	z	P> z	[0.025	0.975]
Intercept	-1.8137	0.405	-4.482	0.000	-2.607	-1.021
pt_width	-0.7694	0.365	-2.106	0.035	-1.485	-0.053
pt_length	-0.4302	0.366	-1.176	0.240	-1.147	0.287
fur_heat_time	0.5227	0.319	1.637	0.102	-0.103	1.148
fur_soak_temp	1.5911	0.480	3.312	0.001	0.649	2.533
rolling_temp	1.3481	0.576	2.342	0.019	0.220	2.476
descaling_count	-0.6708	0.319	-2.106	0.035	-1.295	-0.046

앞서 이진분류모델에서 중요한 변수로 판단된 fur_heat_temp, pt_length를 제외한 모든 변수들의 p-value값이 0.05이하이므로 유의한 변수임을 확인할 수 있다.

1. 모델 별 오차 분석

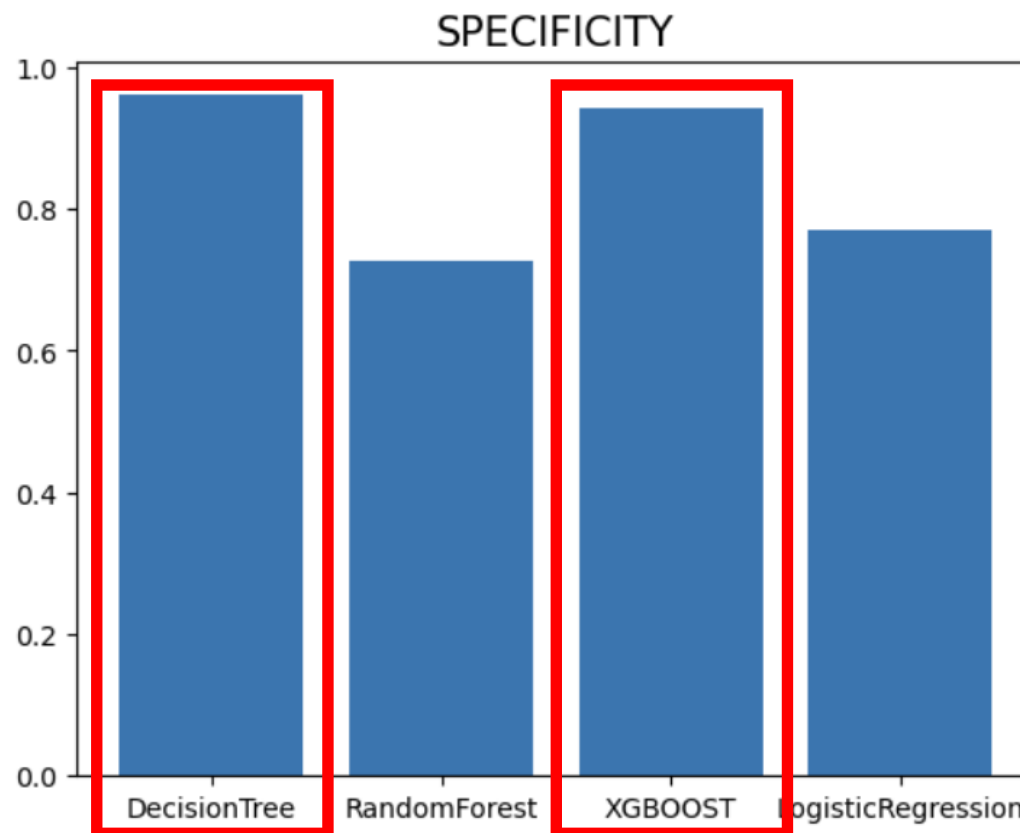
모델	Decision Tree	Random Forest	XGB	Logistic Regression
정확도 (기본 모델 기준)	Training : 0.960% Test : 0.954%	Training : 0.957% Test : 0.954%	Training : 0.958% Test : 0.958%	Training : 0.849% Test : 0.908%

➔ 학습 데이터와 테스트 데이터 모두에서 비슷한 정도의 설명력을 보인다. 따라서 위 표의 전반적인 결과를 토대로 XGB를 최종 모델로 선택



Accuracy, F1_score, Recall, Precision 에서도
XGB가 가장 성능이 우수하다고 나타남

1. 모델 별 오차 분석

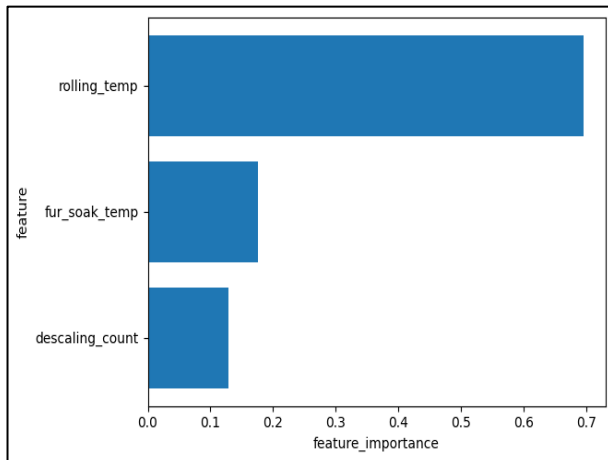


불량 제품을 양품으로 판단하는 오류를 주의해야하므로
특이도를 모델 성능 판단의 기준으로 잡는다.
→ 의사결정나무와 XGBOOST 모델로 불량 분석 및 관리

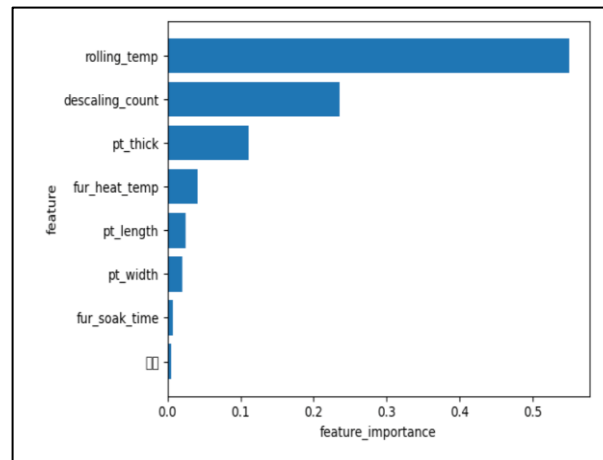
2. 설명 변수의 중요도 결과

〈가격에 영향을 미치는 핵심 인자〉

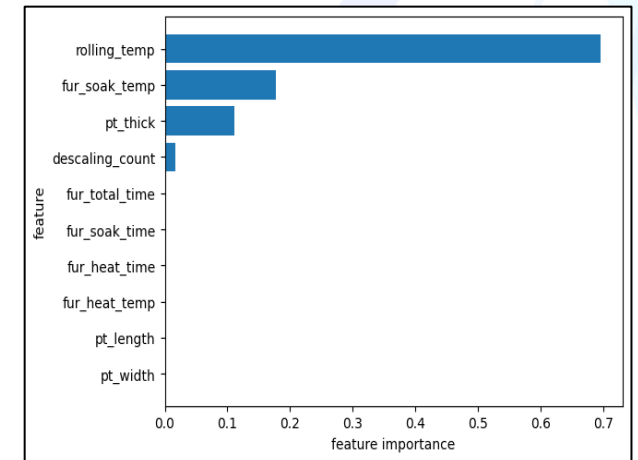
Decision Tree, Random Forest, Gradient Boosting 의 중요도 그래프를 비교해보았다.



Decision Tree



Random Forest



XGB

결론

- 모델별로 순서의 차이가 있지만 압연 온도, 지시 두께, 추출 온도가 스케일에 영향을 준다는 것을 확인할 수 있다.
- 상관관계 분석과 의사결정나무의 결과를 고려했을 때, 압연 온도와 추출 온도가 임계값보다 크고, 지시 두께가 작을수록 불량률이 커진다는 것을 확인할 수 있다.

1. 프로젝트를 통해 얻은 인사이트

후판 제조 과정에서 스케일(scale) 형성은 주로 고온에 의한 산화 반응으로 인해 발생한다. 이 과정에서 스케일의 양을 줄이기 위해 HSB와 descaling 공정을 진행한다. 이러한 공정을 통해, 스케일을 제거함으로써 스케일의 양을 줄일 수 있습니다. 스케일 형성에 가장 큰 영향을 미치는 요소는 **후판의 온도로**, 고온은 금속의 산화 반응을 촉진시켜 **스케일 형성을 가속화한다**.

이는 가열대, 균열대의 소재 온도, 추출 온도, 그리고 특히 **압연 온도**에도 적용된다. 압연 공정 중 금속이 대기와의 직접적인 접촉과 함께 기계적 변형 및 표면 균열로 인해 스케일 형성이 더욱 촉진된다. 따라서, **압연 온도가** 스케일 형성에서 **가장 결정적인 변수임을 시사한다**. 또한, 후판의 두께는 스케일 형성에 영향을 미칠 수 있다. 두께가 두꺼울수록 내부 온도 불균형으로 인해 후판의 재료 균열이 발생할 가능성이 있으며, 이는 내부 스케일 형성을 촉진한다.

후판의 두께와 descaling의 상관 관계가 높은 이유는 후판의 두께가 얇아질수록 압연 공정이 더 많이 필요하고, 이로 인해 스케일이 더 많이 발생할 수 있으며, 그래서 descaling 공정의 횟수가 증가하는 것으로 예측한다. 요약하면, 후판 제조 과정에서 스케일 형성은 주로 고온에 의한 산화 반응으로 인해 발생하며, 이를 최소화하기 위해서는 HSB와 descaling 공정이 중요하다. 가장 중요한 영향 요소는 압연 온도이며, **후판의 두께 또한 스케일 형성에 영향을 미칠 수 있다**.

2. 종합 개선점

최종 개선안

1. 압연온도 한계를 초과하지 않고, HSB는 반드시 적용하고, 공정 온도와 시간을 최적화하여 관리한다.
2. 압연 온도는 오스테나이트 상이 존재하는 범위 내에서 가공성과 scale량 사이에서 최적의 지점을 찾는다.
3. HSB, descaling 공정에서 공정횟수에 따른 scale량을 실험을 통해 구해서 최적의 횟수를 찾는다.
4. 온도를 특정 시점 측정이 아니라 상시모니터링 시스템으로 개선하여 scale 불량 의 핵심 파라미터를 관리한다.