

BigData 종합실습과제

# 압연공정 Scale 불량 예측 및 대응 방안



# Index

## 1 과제 정의

- 과제 추진배경
- 잠재원인 도출

## 2 데이터 수집

- 조사대상 정의
- 데이터 수집
- 데이터 정제

## 3 데이터 분석

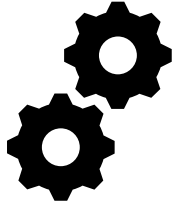
- 분석계획 수립
- 데이터 분석
- 분석결과 정리

## 4 개선안 도출

- 개선안 도출/선정
- 개선안 구체화

# 과제정의 과제추진배경

## 불량률 증가



제철공장의 불량률이 증가  
불량률의 원인으로 압연공정  
Scale 불량 지목  
(불량률 32%)



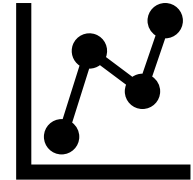
## Scale 문제점



정상부분과 Scale은  
마찰계수가 달라 강판을  
휘어지게 하며,  
설비와의 마찰이 증가해 고  
장의 원인이 된다.



## 원인분석과 개선안 도출



다양한 분류모델을 활용한  
Scale 발생 영향인자 도출  
핵심인자를 통한  
불량 제품 분류 및 개선



# 과제정의 잠재원인 도출

가열대 온도와 Scale 발생 빈도

가열로에서 추출하는 제품 온도

가열로  
관련 인자

압연공정  
과정

사상압연온도

제품 규격에 따른 압연 횟수

고객사에서 요구하는 제품의 규격

(제품 두께, 너비, 길이, 중량 등)

제품  
규격

Scale  
처리 과정

HSB(Hot Scale Braker)

압연 간 Descaling 횟수

### 변수 정보 확인

STEEL_KIND	후판 강종	탄소강 또는 티타늄합금
PT_THICK	후판 지시두께	외주 지시사항
PT_WIDTH	후판 지시폭	
PT_LENGTH	후판 지시길이	
PT_WEIGHT	후판 지시중량	
FUR_NO	가열로 호기	
FUR_NO_ROW	가열로 장입열 구분	
FUR_HZ_TEMP	가열로 가열대 소재온도	제강 공정 관련 변수
FUR_HZ_TIME	가열로 가열대 재로시간	
FUR_SZ_TEMP	가열로 균열대 소재온도	
FUR_SZ_TIME	가열로 균열대 재로시간	
FUR_TIME	가열로 재로시간	
FUR_EXTEMP	가열로 추출온도(계산치)	
ROLLING_TEMP_T5	압연온도(T5)	압연 공정 관련 변수
HSB	HSB 적용여부	Scale 제거 공정
ROLLING_DESCALING	압연 Descaling 횟수	

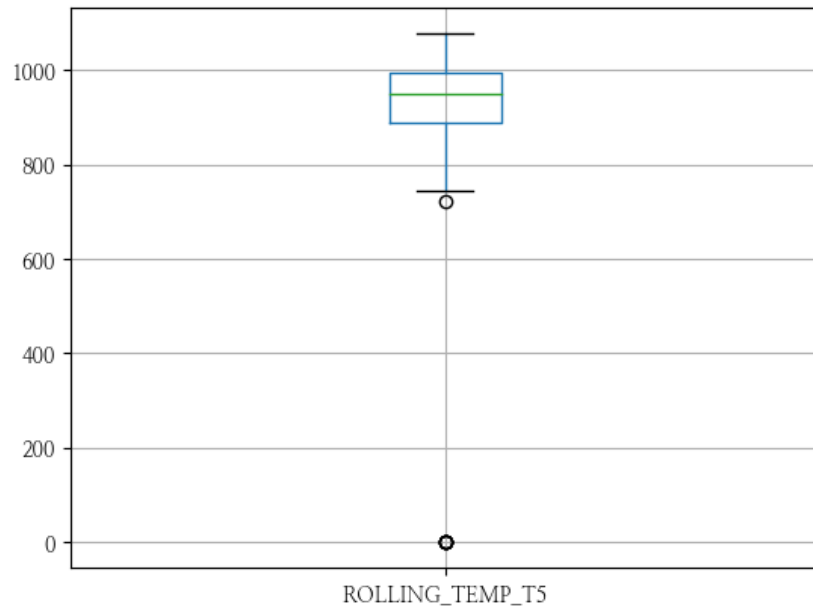
### Describe

	count	mean	std	min	25%	50%	75%	max
PT_THICK	720.0	29.267857	19.152651	12.03	16.09	20.13	40.00	100.34
PT_WIDTH	720.0	2807.127778	517.839690	1800.00	2466.00	2725.00	3125.00	4575.00
PT_LENGTH	720.0	34323.269444	13828.434837	7893.00	22663.00	37380.00	43735.00	54917.00
PT_WEIGHT	720.0	41481.777778	24751.322015	7820.00	17757.50	36120.00	61120.00	115080.00
FUR_HZ_TEMP	720.0	1155.338889	20.866398	1103.00	1139.00	1158.00	1171.00	1206.00
FUR_HZ_TIME	720.0	88.204167	36.248696	43.00	65.00	75.00	101.00	251.00
FUR_SZ_TEMP	720.0	1149.666667	17.571788	1113.00	1132.00	1155.00	1164.00	1185.00
FUR_SZ_TIME	720.0	75.330556	26.086390	35.00	57.75	67.50	85.25	194.00
FUR_TIME	720.0	300.563889	45.390303	221.00	257.00	304.00	342.00	398.00
FUR_EXTEMP	720.0	1139.570833	17.857121	1102.00	1122.75	1144.00	1154.00	1175.00
ROLLING_TEMP_T5	720.0	933.920833	107.863887	0.00	889.75	951.00	994.25	1078.00
ROLLING_DESCALING	720.0	8.319444	1.623252	5.00	6.00	8.00	10.00	10.00

데이터의 전체적인 모습 확인

압연온도변수에서 최소값이 0인 이상치를 식별

## 이상치 식별(1)

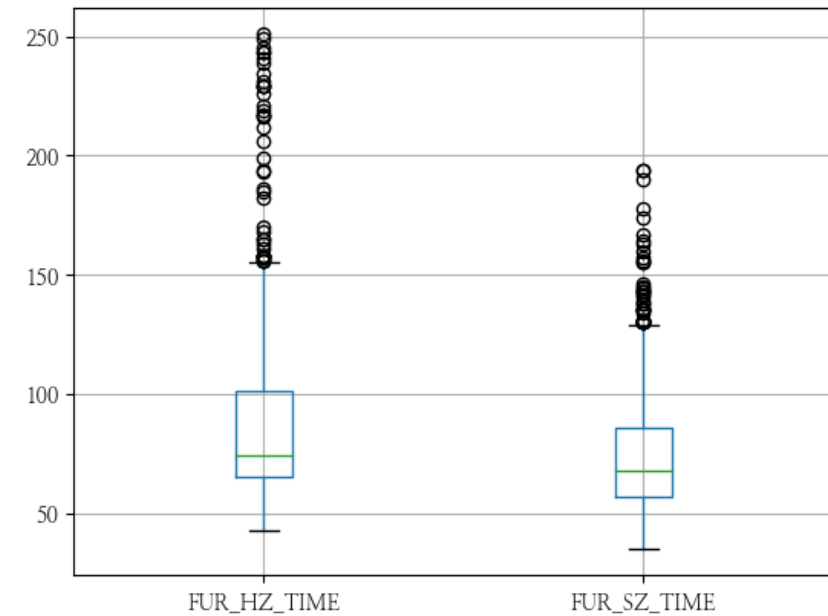


압연온도의 Boxplot

하한선 아래 이상치 식별(7개)

0인 값은 제거, 그 이외의 값은 하한값으로 치환

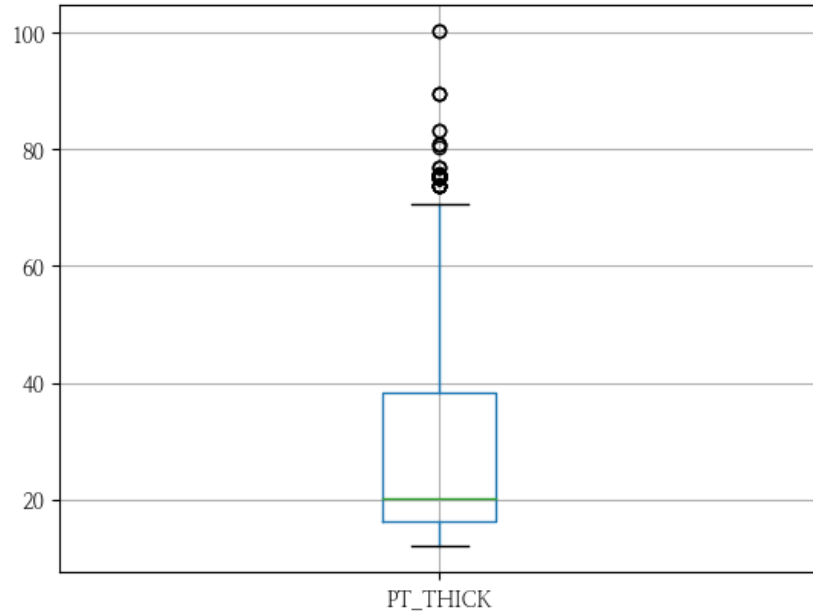
## 이상치 식별(1)



가열로와 균열로 소재시간 Boxplot

상한선 상단의 이상치를 상한선 값으로 치환

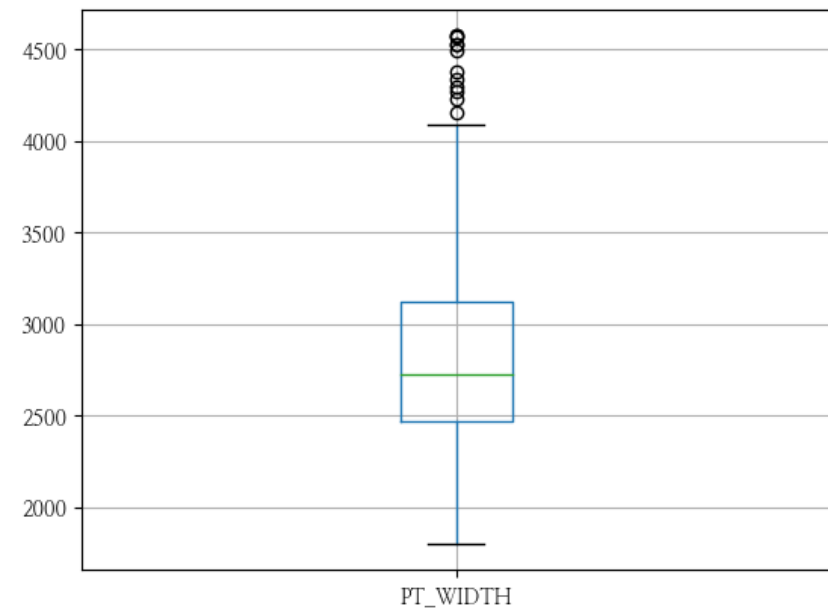
## 이상치 식별



제품 두께의 Boxplot

상한선 상단의 이상치를 상한선 값으로 치환

## 이상치 처리

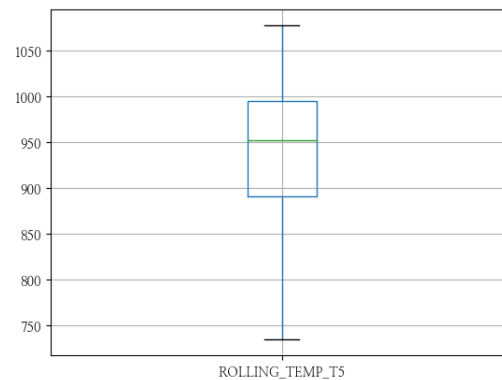
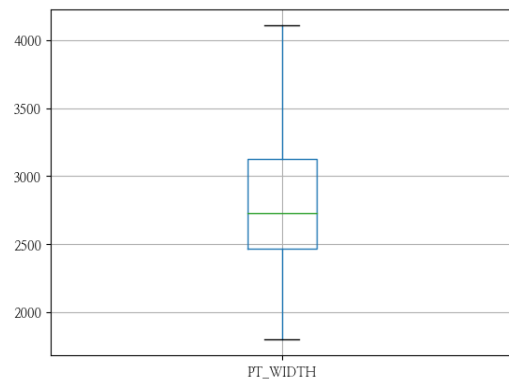
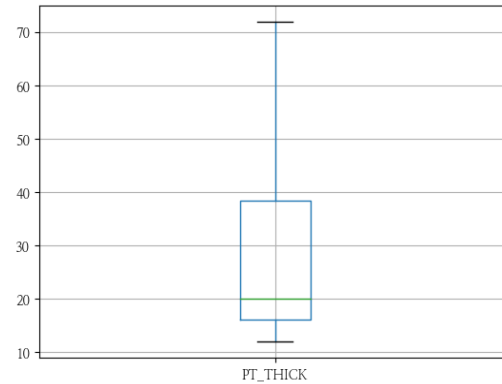
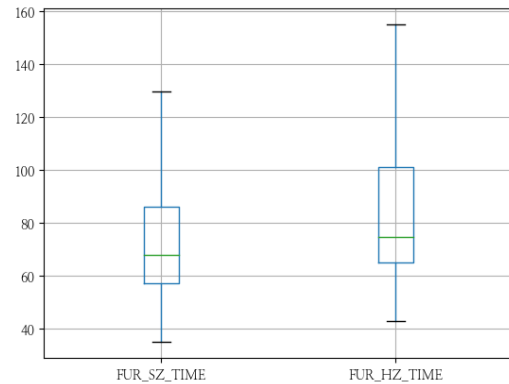


제품 너비의 Boxplot

상한선 상단의 이상치를 상한선 값으로 치환



## 이상치 처리



처리방법 : 중앙값 기준 처리(Clipping)

경계값 외부의 이상치들을 상한값과 하한값으로 치환

상한선 :  $3\text{분위} - 1.5 \times \text{IQR}$

하한선 :  $1\text{분위} + 1.5 \times \text{IQR}$

처리 후 이상치가 상한값 및 하한값으로 치환되어  
Boxplot에 이상치가 보이지 않게 된다.



## 탐색적 분석

기본 설명변수 및  
파생변수의 특징과  
Scale 분포 확인

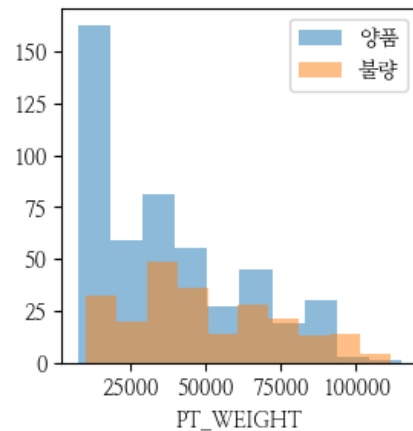
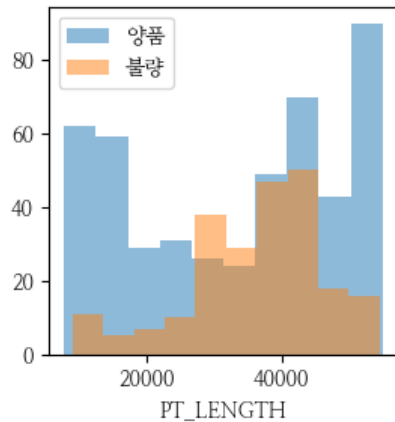
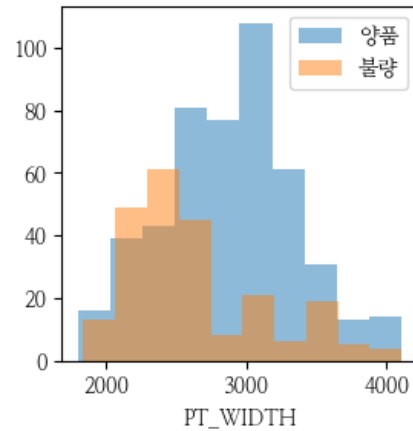
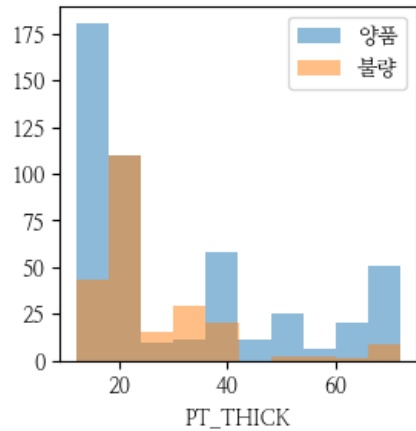
## 추론검정

가열로 가열속도 t-test  
균열로 가열속도 t-test  
압연온도 t-test  
추출온도 t-test

## 모델링

Decision Tree  
SVM  
KNN  
Neural Network

## 제품 규격의 Scale 분포



두께 : 얇을 수록 Scale 비율이 증가하였다.

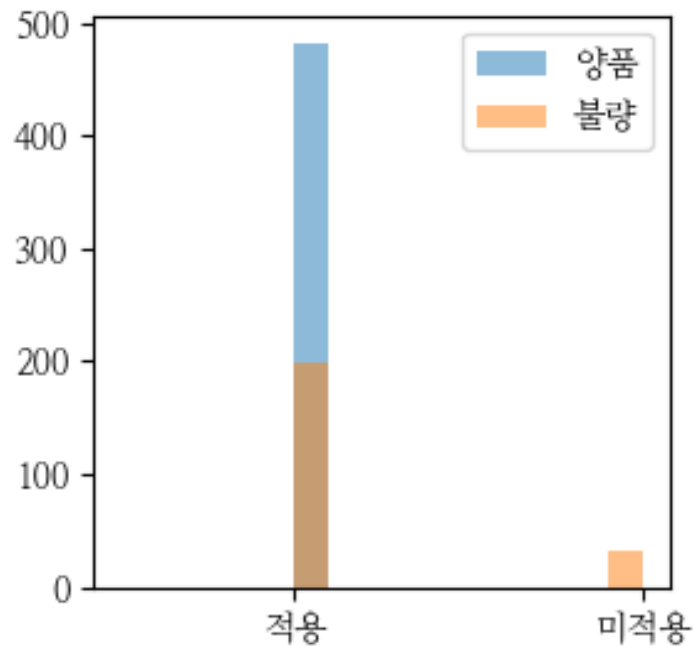
너비 : 좁을 수록 Scale 비율이 증가하였다.

길이 : 3000~4000mm에서 Scale 비율이 증가하였다.

무게 : Scale 비율이 전체적으로 고르게 분포하고 있다.

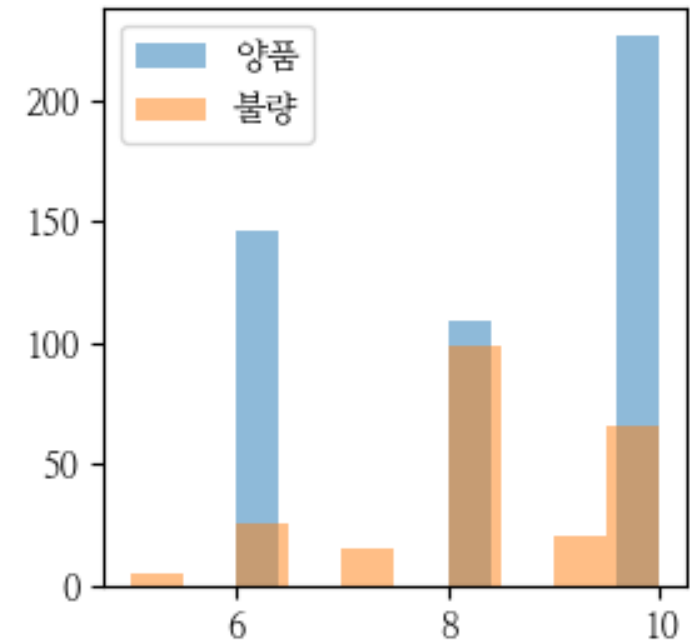
압연 공정에서 제품의 규격에 따라 압연 횟수가 달라지고  
압연 횟수가 증가할 수록 Scale 발생 빈도가 증가한다고  
추정하였음. (추후 가설검정 및 모델링에서 확인할 예정)

## HSB



HSB 적용에 따른 Scale 비율 Boxplot  
HSB를 미적용한 제품은 모두 불량 발생하였다.

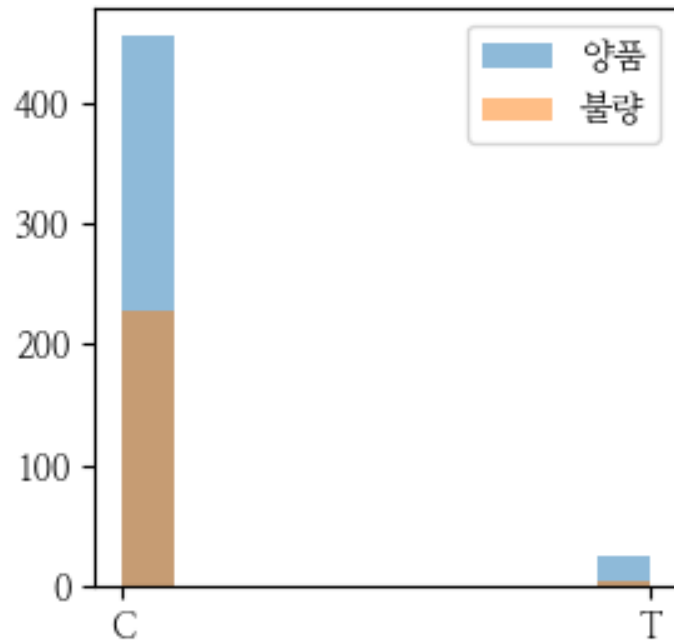
## Descaling



Descal 횟수의 증가에 따른 Scale 비율 Boxplot  
횟수가 증가하여도 Scale 발생 빈도는 감소하지 않음

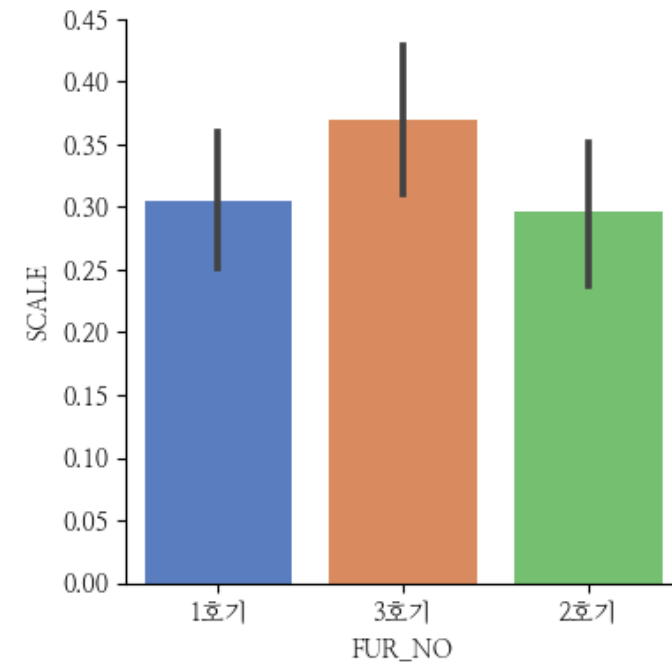


## 제품 강종별 Scale



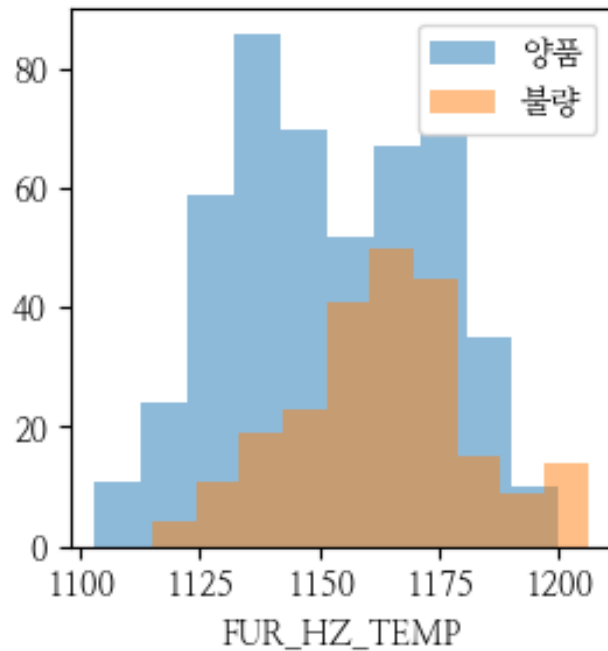
강종별(탄소강, 티타늄 합금) Scale 비율 Boxplot  
탄소강이 티타늄 합금보다 Scale이 더 많이 발생했다.

## 가열로 호기 별 Scale



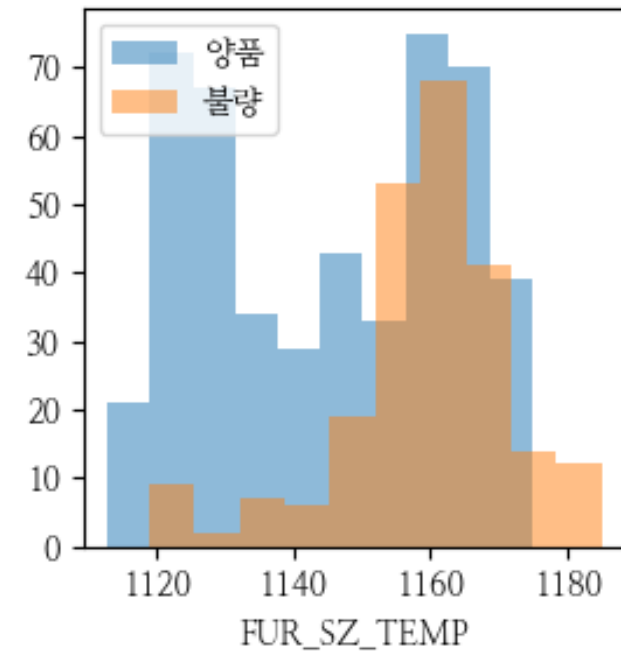
가열로 호기 별 Scale 비율 Boxplot  
호기 별 Scale 차이는 유의미하지 않다.

## 가열로 소재온도와 Scale



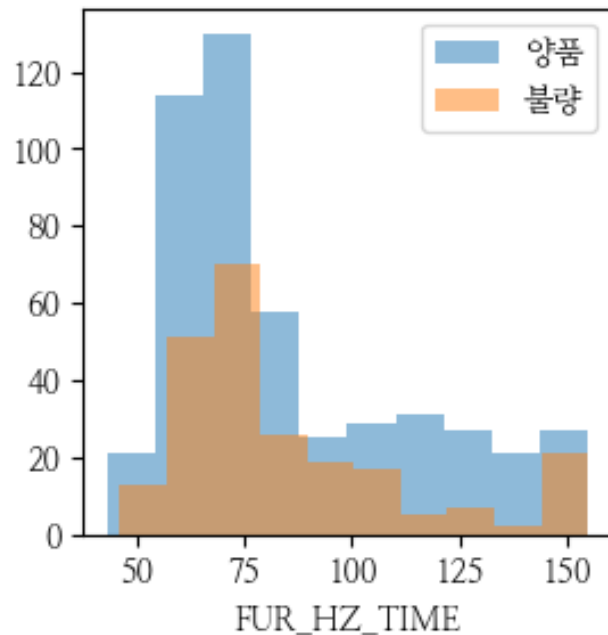
가열로 내부 온도에 따른 Scale 비율 Boxplot  
1150~1175°C에서 Scale 비율이 증가하였다.

## 균열로 소재온도와 Scale



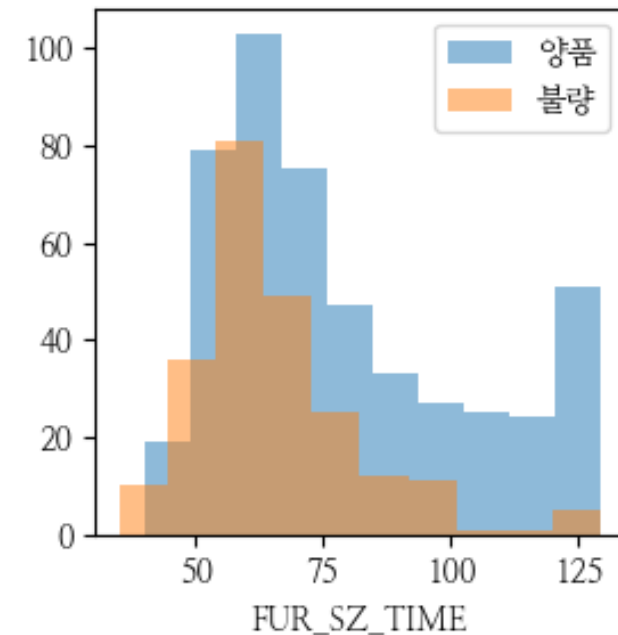
균열로 내부 온도에 따른 Scale 비율 Boxplot  
1150°C 이상에서 Scale 비율이 매우 증가하였다.

## 가열로 재로시간과 Scale



가열로 재로시간에 따른 Scale 비율 Boxplot  
시간이 짧을수록 Scale 비율이 증가하였다.

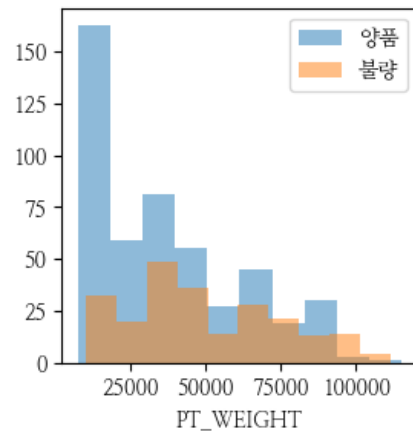
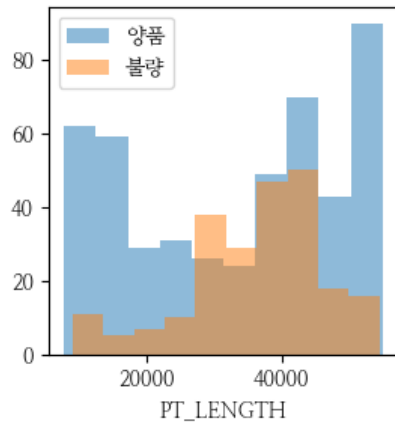
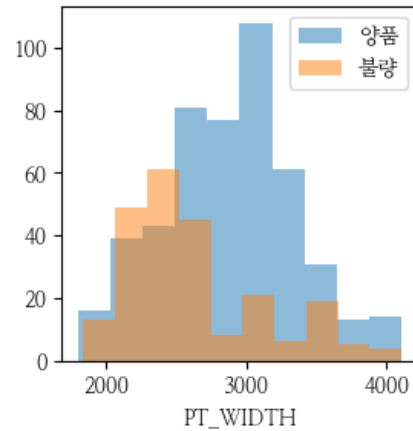
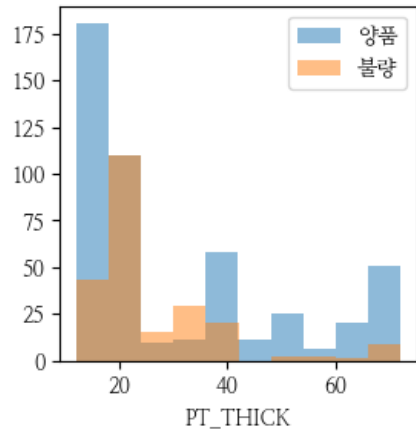
## 균열로 재로시간과 Scale



균열로 재로시간에 따른 Scale 비율 Boxplot  
시간이 짧을수록 Scale 비율이 증가하였다.



## 상관분석 결과 : Scale과 온도의 상관관계



두께 : 얇을 수록 Scale 비율이 증가하였다.

너비 : 좁을 수록 Scale 비율이 증가하였다.

길이 : 3000~4000mm에서 Scale 비율이 증가하였다.

무게 : Scale 비율이 전체적으로 고르게 분포하고 있다.

## 유의미한 결과 정리

가열대 재로시간이 짧을수록  
Scale 발생률이 증가하였고,  
특정 온도범위에서  
Scale 발생률이 증가하였다.

탄소강의 경우, 포함된 Si로  
인하여, 2차 Scale이 발생하여  
불량률이 높게 측정되었다.

재로시간  
및  
온도

제품  
규격

제품 두께가 16mm~40mm  
에서 불량률이 가장 높았고,  
너비가 좁아질수록 불량률이  
증가하였다.

강종별  
차이점

표면처리  
공정의 효과성

HSB 적용시, 불량률은 29%로  
HSB는 효과적이지 못했다.  
Descaling은 횟수를 증가시켜도  
표면 scale 제거효과가 낮았다.

## 가열로와 균열로에서 불량률을 개선할 방안 탐구

가설	P-value	검정결과
가열로 가열 속도와 불량률 사이에는 연관성이 있을 것이다.	0.0001	연관성이 있다.
균열로 가열 속도와 불량률 사이에는 연관성이 있을 것이다.	0.08	연관성이 없다.

Scale이 발생한 제품의 균열로 가열 속도는 18.67로 평균보다 약 1.8°C/s 높지만 연관성이 없다고 검정된 반면, Scale이 발생한 제품의 가열로 가열속도는 14.74로 평균보다 약 0.1°C/s 높지만 연관성이 검정되었다.

**가열로**는 슬라브의 **표면온도 상승이 목적**이며, **균열로**는 **일정하게 온도를 유지**시키는 것이 목적이다.

가열온도가 **너무 높아지면** 탄소강의 경우, 표면에서 Si와 Fe가 반응하여 **red Scale이 발생**하게 되기 때문에 이는 급격한 온도상승이 일어나는 **가열로**에서는 **적절한 온도에서 충분한 예열**이 필요한 것으로 판단된다.



## 압연공정에서 불량률을 개선할 방안 탐구

가설	P-value	검정결과
압연온도와 불량률 사이에는 연관성이 있을 것이다.	2.54 e-13	연관성이 매우 크다.
추출온도와 불량률 사이에는 연관성이 있을 것이다.	3.19 e-7	연관성이 매우 크다.
추출온도와 압연온도의 차이는 불량률에 영향을 미칠 것이다.	7.64 e-7	연관성이 매우 크다.

Scale이 발생한 제품의 압연온도는 평균보다 약  $0.02^{\circ}\text{C}$  낮게, 추출온도는 평균보다 약  $10^{\circ}\text{C}$  높고 측정되었다. 압연공정에서 적절한 수준보다 높은 온도의 탄소강 표면의 경우, 표면에서 Si와 Fe가 반응하여 red Scale이 발생하고, 점도가 높은 화합물이 불규칙적으로 형성되어 추후 Descale 공정에서도 쉽게 제거되지 않는다. 충분히 예열된 슬라브를 적절한 온도에서 신속한 압연공정으로 처리해야 할 것이다.

## 제품 규격별 불량률을 개선할 방안 탐구

가설	P-value	검정결과
제품 두께는 불량률에 영향을 미칠 것이다.	1.46 e-15	연관성이 매우 크다.
제품 너비는 불량률에 영향을 미칠 것이다.	2.04 e-14	연관성이 매우 크다.

4구간으로 구분된 제품 두께에서 불량률이 가장 높은 곳은 20mm~40mm로 51%가 불량으로 측정되었다.

4구간으로 구분된 제품 너비에서 불량률이 가장 높은 곳은 2466mm보다 작은 제품으로 51%가 불량으로 측정되었다.

압연과정에서 발생하고 성장하는 2차 Scale은 압연 Rolling 횟수가 적어 제품이 두꺼워지고 좁아질수록 발생빈도가 증가하는 것으로 추정된다. 두껍고 좁은 제품을 생산시에는 추출온도 조절에 유의하여 최적의 압연온도에서 공정이 이루어 질 수 있도록 해야 할 것이다.

## Accuracy

Accuracy: 0.976

Confusion matrix:

```
[[186  0]
 [ 7  93]]
```

	precision	recall	f1-score	support
0	0.964	1.000	0.982	186
1	1.000	0.930	0.964	100
accuracy			0.976	286
macro avg	0.982	0.965	0.973	286
weighted avg	0.976	0.976	0.975	286

양품 : 0, 불량 : 1

Precision : 100%

Recall : 93.0%

## Hyperparameter

Min\_samples\_leaf : 8

Min\_samples\_split : 20

Max\_depth : 10

## Assessment

Precision과 recall값이 높아

영향인자 분석 모델로 적합하다고 판단함



## Accuracy

Accuracy: 0.885

Confusion matrix:

```
[[185  1]
 [ 32 68]]
```

	precision	recall	f1-score	support
0	0.853	0.995	0.918	186
1	0.986	0.680	0.805	100
accuracy			0.885	286
macro avg	0.919	0.837	0.861	286
weighted avg	0.899	0.885	0.878	286

양품 : 0, 불량 : 1

Precision : 98.6%

Recall : 68.0%%

## Hyperparameter

Min\_samples\_leaf : 8

Min\_samples\_split : 20

Max\_depth : 8

n\_estimators : 120

## Assessment

Precision은 0.986으로 높지만, recall이 0.680으로 낮게 측정되어, 실제 모델로는 적합하지 않음.

불량을 양품으로 잘못 측정할 확률이 0.320으로 현재 문제를 개선하지 못할 것으로 판단함.

# 데이터 분석 : 모델링(Gradient Boosting)

## Accuracy

Confusion matrix:

```
[[186  0]
 [ 3  97]]
```

	precision	recall	f1-score	support
0	0.984	1.000	0.992	186
1	1.000	0.970	0.985	100
accuracy			0.990	286
macro avg	0.992	0.985	0.988	286
weighted avg	0.990	0.990	0.989	286

양품 : 0, 불량 : 1

Precision : 100%

Recall : 97.0%%

## Hyperparameter

Min\_samples\_leaf : 7

Min\_samples\_split : 30

Max\_depth : 10

n\_estimators : 70

Learning\_rate : 0.4

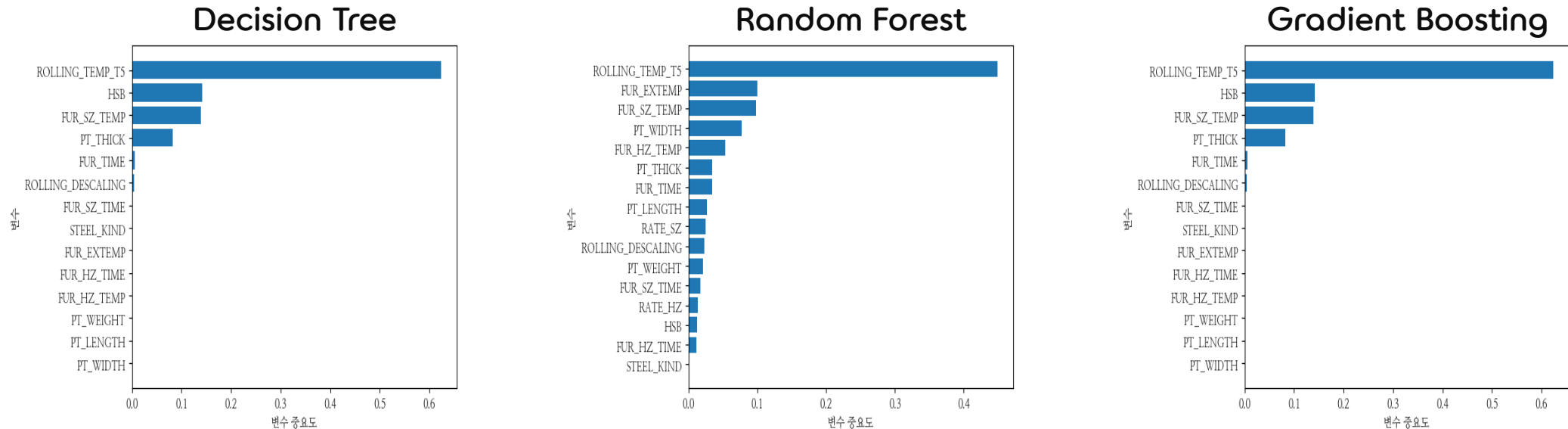
## Assessment

Train Accuracy가 1.000으로 과대적합적이지만,

Test Accuracy도 0.990으로 높게 형성되었다.

영향인자 분석 모델로 적합하다고 판단함.

## Feature Importance



Tree 계열 알고리즘에서 공통적으로 영향이 높은 인자는 압연온도로 측정되었다. 이는 EDA분석과 일치하여 해당 변수를 조절하면 공정과정의 Scale을 감소시킬 수 있다고 판단하였다. 또한 HSB 적용 여부가 큰 영향을 미치며, 제품의 규격(두께, 너비)과 가열로에서의 온도가 중요한 영향인자로 판단되었다.



## Accuracy

Confusion matrix :

```
[[180  6]
 [ 21 79]]
```

	precision	recall	f1-score	support
0	0.896	0.968	0.930	186
1	0.929	0.790	0.854	100
accuracy			0.906	286
macro avg	0.912	0.879	0.892	286
weighted avg	0.907	0.906	0.904	286

양품 : 0, 불량 : 1

Precision : 92.9%

Recall : 79.0%

## Hyperparameter

Cost : 1.7

Gamma : 0.05

## Assessment

Precision은 높지만, Recall이 낮기 때문에  
실제 공정과정에서 불량을 예측할 수 있는 모델로는  
적합하지 않다고 판단함.

## Accuracy

```
Accuracy on training set: 1.000
Accuracy on test set: 0.895
Confusion matrix:
[[174  12]
 [ 18  82]]
```

	precision	recall	f1-score	support
0	0.906	0.935	0.921	186
1	0.872	0.820	0.845	100
accuracy			0.895	286
macro avg	0.889	0.878	0.883	286
weighted avg	0.894	0.895	0.894	286

양품 : 0, 불량 : 1

Precision : 87.2%

Recall : 82.0%

## Hyperparameter

Batch Size : 50

Solver : L-BFGS

Activation : ReLu

Hidden Layer : 21

## Assessment

Precision과 Recall은 Gradient Boosting보다 낮게 측정되었기 때문에 실제 활용성을 떨어진다고 판단하였음.

## Accuracy

```
Accuracy on train set : 1.000
Accuracy on test set : 0.881
Confusion matrix :
[[178   8]
 [ 26  74]]
```

	precision	recall	f1-score	support
0	0.873	0.957	0.913	186
1	0.902	0.740	0.813	100
accuracy			0.881	286
macro avg	0.887	0.848	0.863	286
weighted avg	0.883	0.881	0.878	286

양품 : 0, 불량 : 1

Precision : 90.2%

Recall : 74.0%

## Hyperparameter

이웃의 수(k) : 10

거리계산방법 : Euclidean(L2 norm)

가중치 함수 : Distance(거리의 역수)

## Assessment

Precision은 0.9이상이지만, recall이 낮기 때문에 실제 불량 예측에서는 활용성이 떨어진다고 판단함.



## HSB와 Descaling

Descaling이 증가하더라도 **액상페야라이트** ( $\text{Fe}_2\text{SiO}_4$ )는 강판과의 **밀착성이 높아** 강한 수압으로도 제거되지 않는다. 따라서 HSB 과정은 필수적이며, 처리 과정에서 페야라이트( $\text{Fe}_2\text{SiO}_4$ )의 **녹는점 이상**에서 실시되어야 발생하는 2차 Scale 피해를 줄일 수 있을 것이다.

## 합금에 포함되는 성분

액상 페야라이트가 압연중 강판에 치입되면서 발생하는 red scale은 **탄소강에 포함된 Si**가 가열온도  $1173^\circ\text{C}$ 이상인 조건에서 강판 표면의 **철과 산화철 그리고 Si가 반응**하여 생성되는 2차화합물로 인해 발생하며, 이로 인하여 **탄소강의 불량률이 티타늄 합금보다 높다.**

## 제품의 규격

---

제품의 규격과 불량률은 관련성을 보였다.

두께 20mm~40mm 근처와, 너비 2500mm

근처에서 불량률이 증가하였다.

왜 규격에 따라 불량률이 증가하는가

## 추출온도 및 압연온도

---

압연온도는 조절가능한 변수가 아니다.

압연공정에서 온도조절장치가 없기 때문에

추출온도를 조절하여 압연공정 시에 적절한

온도를 유지할 수 있도록 조정

온도가 높아지면 불량률이 떨어지는가?

## 최종 모델에 사용할 인자

압연공정에서 온도조절장치가 없기 때문에 압연온도는 조절가능한 변수가 아니다.

따라서 **추출온도**와 **제품 규격**을 이용하여

**다중선형회귀모델**을 제작하여 **압출온도를 예측**하고

예측한 압출온도를 제품 규격, 가열대의 온도 등과

함께 **Gradient Boosting**으로 분석하여

**불량제품을 분류**할 것이다.

## 최종 모델 평가

주어진 온도와 제품 규격변수로 압연온도를 예측한

**회귀모델의 정확성**은 **60.5%**로 낮게 측정되었다.

이로 인하여 **GB 모델의 recall** 또한 **0.690**으로

낮게 측정되었다. 추가적인 데이터 수집을 통해

가열대의 압력, 추출 후 압연공정 전까지의

이동시간 등, **추가적 영향인자**가 주어진다면,

회귀모델 및 분류모델의 **예측성은 증가**할 것이다.



## Gradient Boosting

Accuracy: 0.892

Confusion matrix:

```
[[186  0]
 [ 31 69]]
```

	precision	recall	f1-score	support
0	0.857	1.000	0.923	186
1	1.000	0.690	0.817	100
accuracy			0.892	286
macro avg	0.929	0.845	0.870	286
weighted avg	0.907	0.892	0.886	286

## 압연온도 추정 Regression

OLS Regression Results						
Dep. Variable:	ROLLING_TEMP_TS		R-squared:	0.611		
Model:	OLS		Adj. R-squared:	0.605		
Method:	Least Squares		F-statistic:	109.8		
Date:	Tue, 08 Nov 2022		Prob (F-statistic):	7.75e-83		
Time:	12:38:18		Log-Likelihood:	-2192.9		
No. Observations:	427		AIC:	4400.		
Df Residuals:	420		BIC:	4428.		
Df Model:	6					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
Intercept	-1007.5095	193.418	-5.209	0.000	-1387.698	-627.321
PT_THICK	-2.1476	0.270	-7.947	0.000	-2.679	-1.616
PT_WIDTH	-0.0456	0.005	-9.972	0.000	-0.055	-0.037
PT_LENGTH	-0.0020	0.000	-6.266	0.000	-0.003	-0.001
PT_WEIGHT	0.0005	9.61e-05	4.856	0.000	0.000	0.001
FUR_SZ_TEMP	2.7766	0.234	11.871	0.000	2.317	3.236
FUR_HZ_TEMP	-0.8700	0.161	-5.390	0.000	-1.187	-0.553
Omnibus:	21.040	Durbin-Watson:	2.050			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	56.695			
Skew:	-0.100	Prob(JB):	4.89e-13			
Kurtosis:	4.774	Cond. No.	5.69e+06			

## 세밀한 온도 조절

---

가열로에서는 슬라브의 표면온도가  $1173^{\circ}\text{C}$ 를 넘지 않도록 온도와 압력을 조절하여 Scale의 발생률을 낮출 수 있다.

## 분류모델을 활용한 불량 예측

---

추출온도를 이용하여 불량모델을 예측하고  
공정 과정에서 추출온도를 실시간으로 측정하여  
불량발생률이 높은 온도가 되면 자동으로  
가열대와 압연 공정이 조절되도록 시스템 구축



The background of the slide features a close-up, high-contrast photograph of several interlocking metal gears. The gears are made of a dark, polished metal, likely steel, and their teeth are sharp and well-defined. The lighting creates strong highlights and shadows, emphasizing the mechanical texture and the circular patterns of the gears. The composition is a partial view, showing the edges and teeth of the gears as they mesh together.

**감사합니다**