

00 공장의 Scale 불량 인자 확인 및 분류

A반 강연주

00 공장의 Scale 불량 발생 증가의 원인 파악 및 개선 기회 도출

- 00공장의 압연 공정에서 scale 불량 급증으로 인한 불량률 증가
- 압연 공정 결과 데이터를 바탕으로 scale 불량률 증가의 원인 파악 및 개선 기회 도출

잠재적 인자 선정

Scale 발생	가열로 가열대 온도	가열로 균열대 온도	가열로 추출 온도	Hot Scale Breaker	사상 압연 온도	압연 간 Descaling 횟수	판두께
없음 ↕ 발생	저 ↕ 고	저 ↕ 고	저 ↕ 고	적용 ↕ 미적용	저 ↕ 고	증가 ↕ 감소	후 ↕ 박

- Scale 불량에 영향을 미치는 잠재 인자 선정 및
방향성 확인
- 도출된 잠재 인자와 분석 결과 비교 예정

변수 탐색

- 목표 변수

- SCALE : Scale 불량여부

- 설명 변수

1. 판의 특성

1) 화학적 특성

- STEEL_KIND : 강종

2) 물리적 특성

- PT_THK : plate의 두께
- PT_WIDTH : plate의 폭
- PT_LTH : plate의 길이
- PT_WGT : plate의 중량

2. 기타 특성

- ROLLING_DATE : 작업시각
- SPEC : 제품 규격
- WORK_GR : 작업조

3. 가열로 특성

- FUR_NO : 가열로 호기
- FUR_NO_ROW : 가열로 작업 순번
- FUR_HZ_TEMP : 가열로 가열대 온도
- FUR_HZ_TIME : 가열로 가열대 시간
- FUR_SZ_TEMP : 가열로 균열대 온도
- FUR_SZ_TIME : 가열로 균열대 시간
- FUR_TIME : 가열로 시간
- FUR_EXTEMP : 추출 온도

4. 압연 특성

- ROLLING_TEMP_T5 : 압연 온도
- HSB(Hot_Scale_Breaker):
강편의 표면에 산화작용으로 생긴 스케일을 제거하는 기구
- ROLLING_DESCALING : 압연 중 descaling 횟수

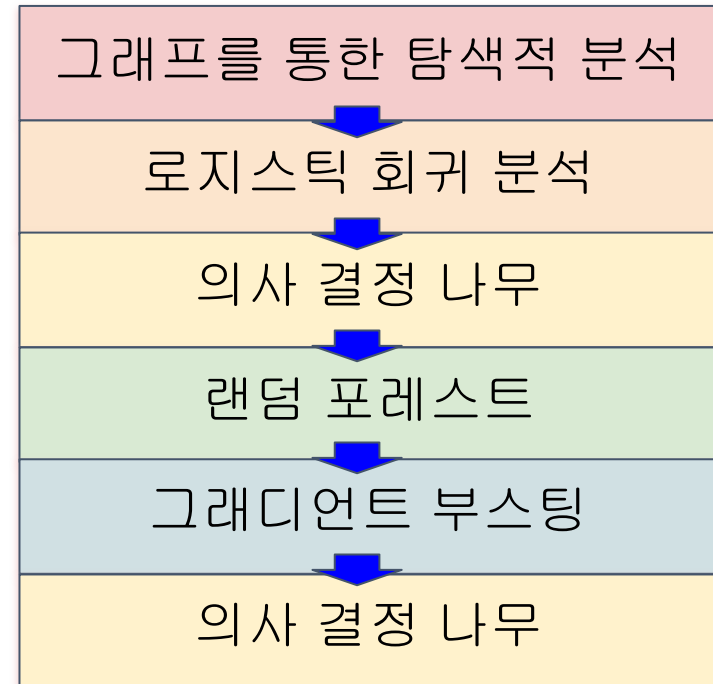
데이터 분석 방향 및 기법

랜덤 포레스트

● 분석 방향

- 다양한 분류 모델을 통해 Scale 불량 발생에 영향을 주는 인자 도출
- 도출된 영향인자와 잠재 요인과의 일치, 방향성 확인
- 분류 모델의 특징을 이해 및 모델의 성능 개선 방안 도출

● 분석 기법



데이터 전처리

1. 결측치 처리

```
# 결측치 처리
df_raw.isnull().sum(axis=0)
```

PLATE_NO	0
ROLLING_DATE	0
SCALE	0
SPEC	0
STEEL_KIND	0
PT_THK	0
PT_WIDTH	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_EXTMP	0
ROLLING_TEMP_T5	0
HSB	0
ROLLING_DESCALING	0
WORK_GR	0
dtype: int64	

2. 타입 분석

```
# 변수별 타입 분석
df_raw.dtypes
```

PLATE_NO	object
ROLLING_DATE	object
SCALE	object
SPEC	object
STEEL_KIND	object
PT_THK	float64
PT_WIDTH	int64
PT_LTH	int64
PT_WGT	int64
FUR_NO	object
FUR_NO_ROW	int64
FUR_HZ_TEMP	int64
FUR_HZ_TIME	int64
FUR_SZ_TEMP	int64
FUR_SZ_TIME	int64
FUR_TIME	int64
FUR_EXTMP	int64
ROLLING_TEMP_T5	int64
HSB	object
ROLLING_DESCALING	int64
WORK_GR	object
dtype: object	

3. Scale 양품 → 0, 불량 → 1 변환

```
# scale 양품은 0, 불량은 1로 바꾸기
df_raw['SCALE'] = df_raw['SCALE'].map(lambda x : 0 if x == "양품" else 1)
```

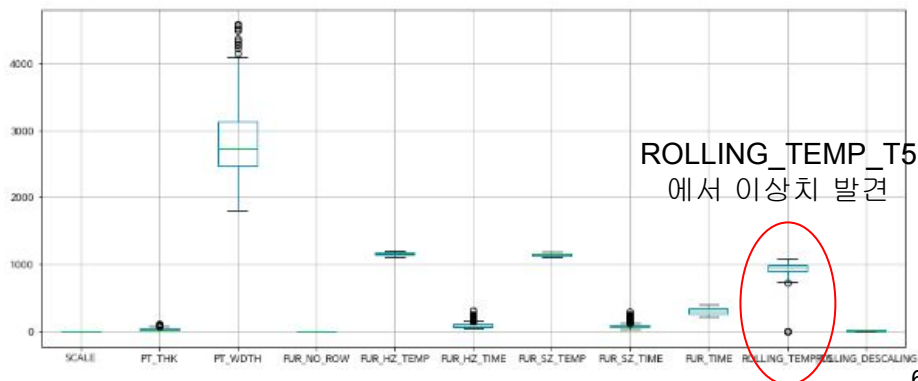
4. 일부 칼럼 제거 - 고유명, 중복, 시간데이터 불필요하다고 판단

```
# 불필요 변수 제거
df_raw.drop(["PLATE_NO", "SPEC", "FUR_EXTMP", "ROLLING_DATE"], axis=1, inplace=True)
```

5. 연속형, 범주형 변수 구분

```
# 연속형 변수, 범주형 변수 구분
df_raw_numeric=df_raw.select_dtypes(exclude='object')
df_raw_object=df_raw.select_dtypes(include='object')
```

6. 이상치 제거

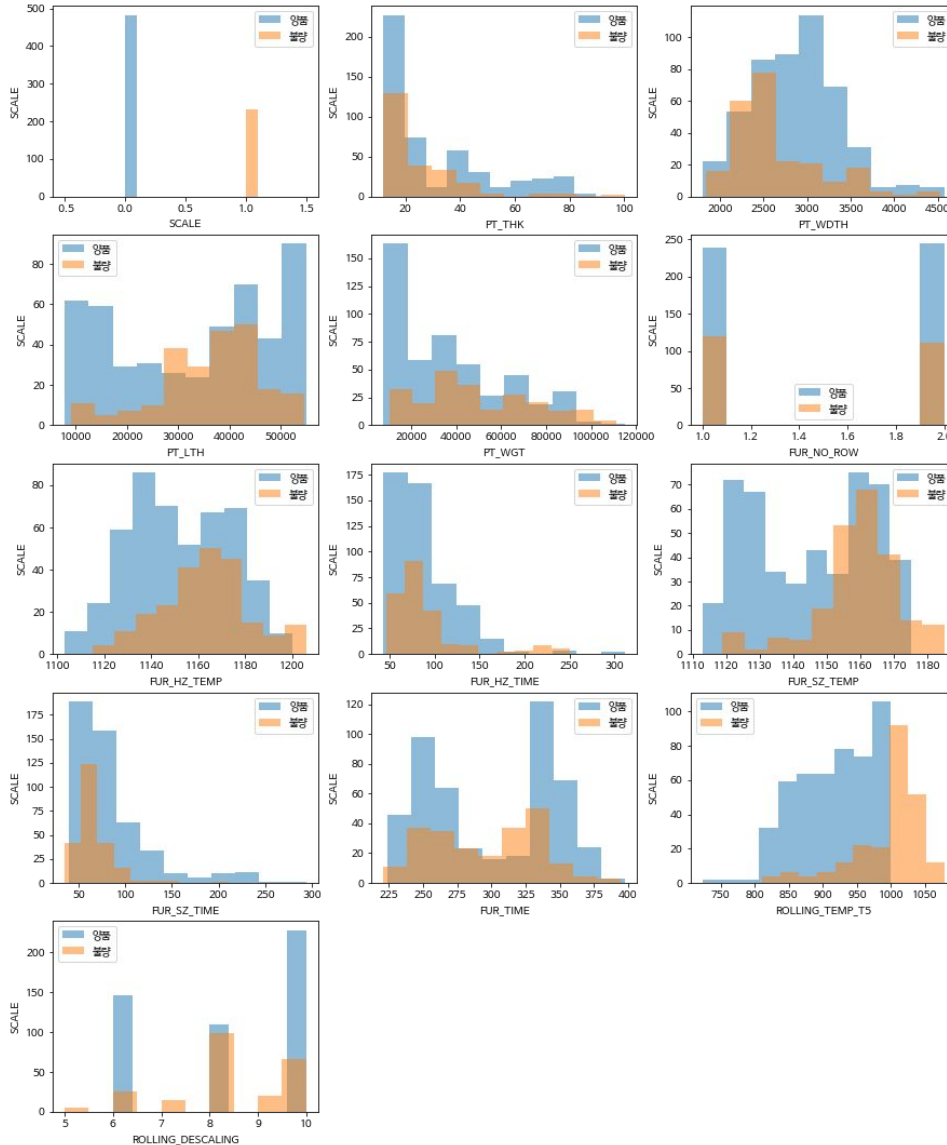
Box plot
그리기

```
# 이상치 제거
df_raw_numeric.boxplot(figsize=(15,6))
```

이상치 제거

```
# 이상치 제거
df_raw.sort_values(by=["ROLLING_TEMP_T5"], axis=0, ascending=True)[0:20]
```

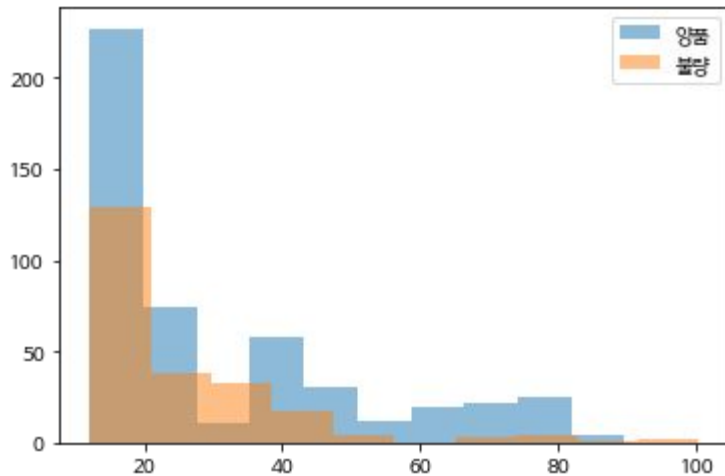
Plate 의 특성에 따른 양품/불량 경향성 by Histogram



- 가열로 가열대 온도 1150도 이상에서 불량률 급격히 증가
- 가열로 균열대 온도 1145도 이상에서 불량률 급격히 증가
- 압연온도 1000도 이상에서는 불량률 100%

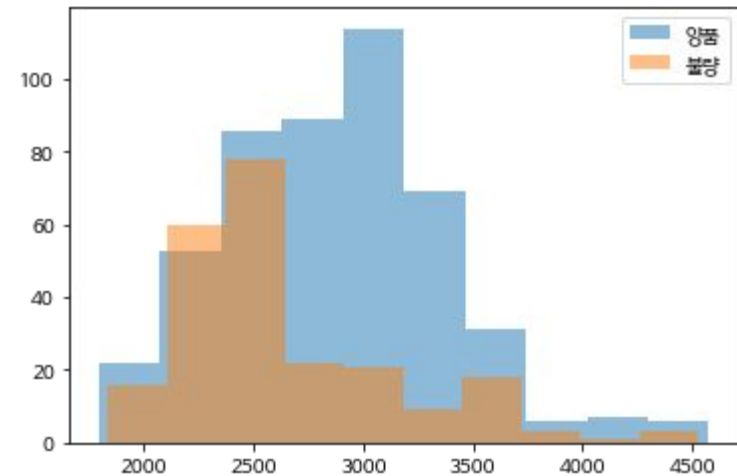
Plate 의 특성에 따른 양품/불량 비율 by Histogram

- Plate의 두께



⇒ $50 < \text{Plate 두께} < 80$, 낮은 불량률

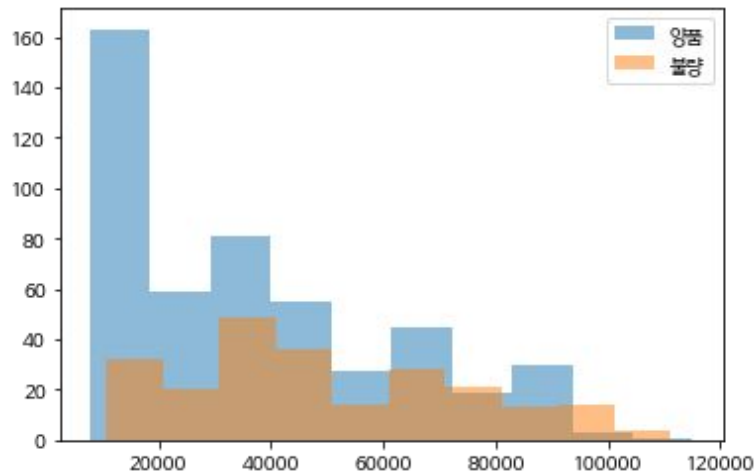
- Plate 의 폭



⇒ $2600 < \text{Plate의 폭} < 3450$, 낮은 불량률

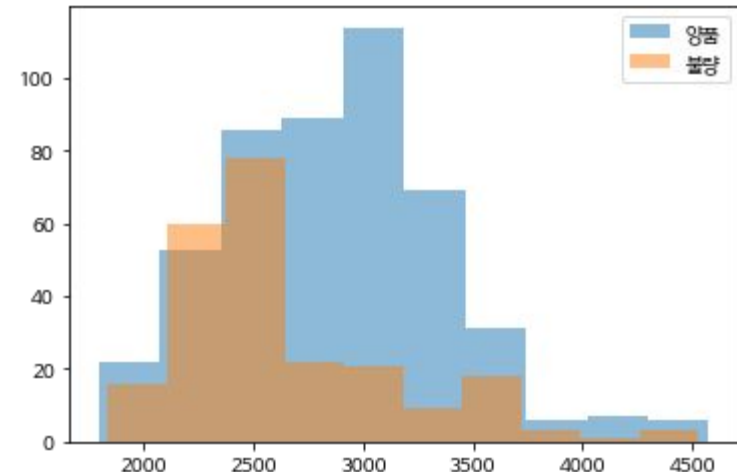
Plate의 특성에 따른 양품/불량 비율 by Histogram

● Plate 의 길이



⇒ Plate 길이 < 2800 or
Plate 길이 > 4500 ,
낮은 불량률

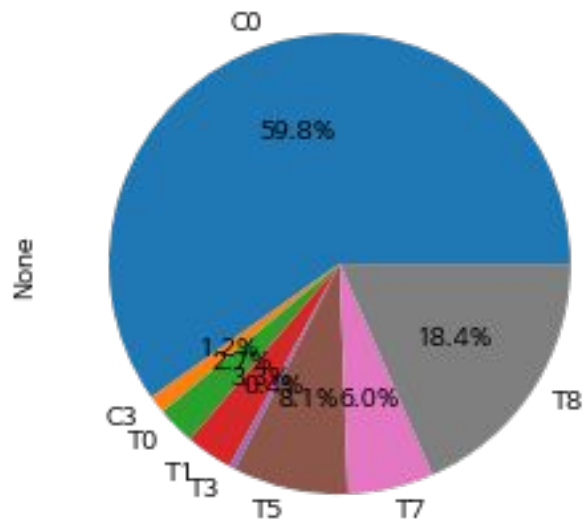
● Plate 의 중량



⇒ Plate의 중량 < 20000,
낮은 불량률

양품 강종의 비율 by Pie Chart and Bar Chart

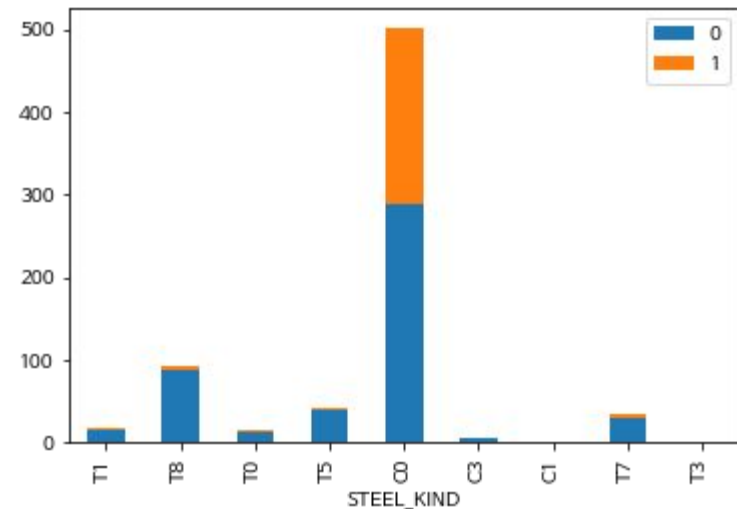
- 양품인 강종의 비율



⇒ C0 가 가장 높은 양품 비율

- ▶▶ 강종 중 C0의 비중이 가장 높음
- ▶▶ 각 강종 중 C0의 불량률이 가장 높음

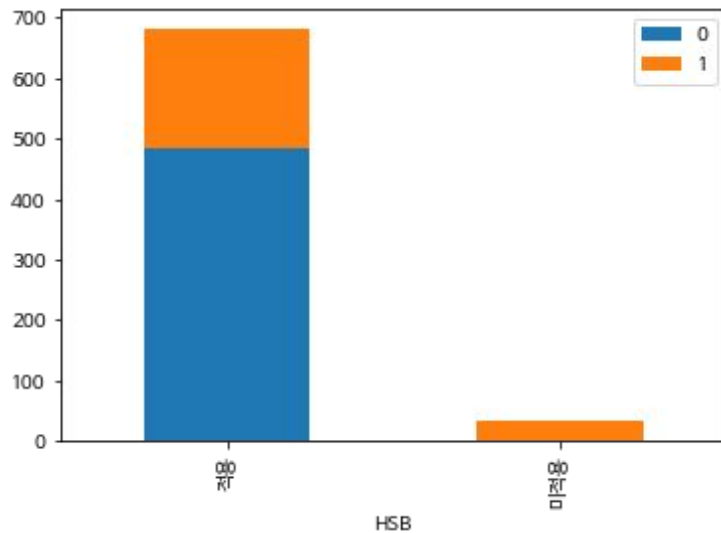
- 각 강종별 양품 비율



⇒ C0가 불량률이 가장 큼

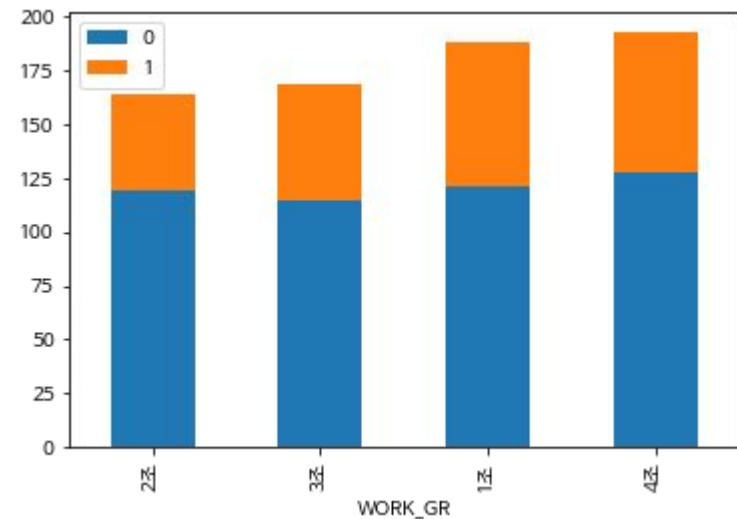
HSB 적용 여부/ 작업조에 따른 불량률 평가

- HSB



⇒ HSB 미적용시 불량률 100%

- 작업조에 따른 분류



⇒ 2,3 조에 비해 1,4조에서 불량률이 다소 큼

가열로에 따른 불량률 평가

- 가열로 호기에 따른 분류

FUR_NO SCALE	1호기	2호기	3호기
0	166	166	151
1	73	70	88

FUR_NO SCALE	1호기	2호기	3호기
0	0.695	0.703	0.632
1	0.305	0.297	0.368

⇒ 각 가열로의 불량률
 3호기(36.8%) > 1호기(30.5%)
 > 2호기(29.7%)

- 가열로 작업 순번에 따른 분류

FUR_NO_ROW SCALE	1	2
0	239	244
1	120	111

FUR_NO_ROW SCALE	1	2
0	0.666	0.687
1	0.334	0.313

⇒ 각 순번의 불량률
 1번(33.1%) > 2번(31%)

로지스틱 선형회귀

Warning: Maximum number of iterations has been exceeded.
 Current function value: 0.257840
 Iterations: 35
 Function evaluations: 51
 Gradient evaluations: 40

Logit Regression Results

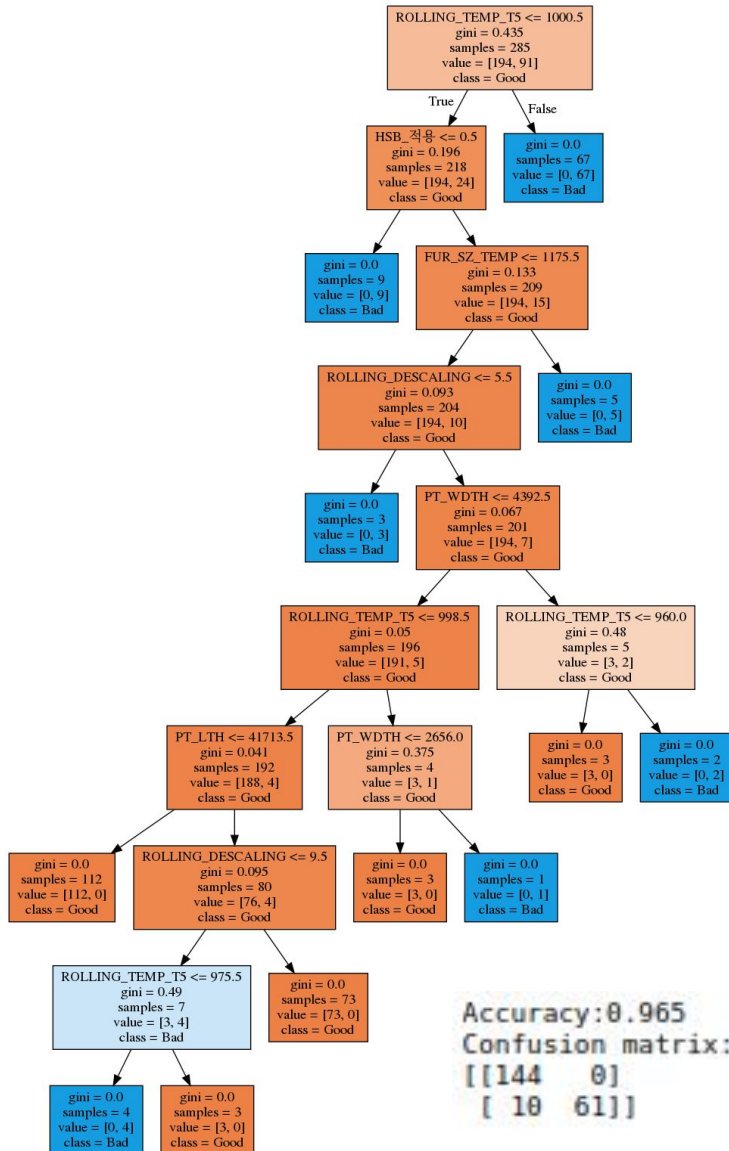
Dep. Variable:	SCALE	No. Observations:	285
Model:	Logit	Df Residuals:	260
Method:	MLE	Df Model:	24
Date:	Tue, 24 Nov 2020	Pseudo R-squ.:	0.5883
Time:	19:57:39	Log-Likelihood:	-73.484
Converged:	False	LL-Null:	-178.51
Covariance Type:	nonrobust	LLR p-value:	1.174e-31

	coef	std err	z	P> z	[0.025	0.975]
Intercept	-0.2550	38.733	-0.007	0.995	-76.170	75.660
C(STEEL_KIND)[T.C3]	-0.1095	4.858	-0.023	0.982	-9.631	9.412
C(STEEL_KIND)[T.T0]	-1.1989	2.180	-0.550	0.582	-5.472	3.074
C(STEEL_KIND)[T.T1]	-1.3454	2.097	-0.642	0.521	-5.456	2.765
C(STEEL_KIND)[T.T5]	-1.9929	3.194	-0.624	0.533	-8.252	4.267
C(STEEL_KIND)[T.T7]	-0.4482	3.292	-0.136	0.892	-6.900	6.004
C(STEEL_KIND)[T.T8]	1.2719	2.383	0.534	0.594	-3.400	5.943
C(FUR_NO)[T.2호기]	0.0929	0.553	0.168	0.867	-0.991	1.177
C(FUR_NO)[T.3호기]	1.1418	0.599	1.908	0.056	-0.031	2.315
C(HSB)[T.적응]	-9.4529	2.865	-3.300	0.001	-15.068	-3.838
C(WORK_GR)[T.2조]	-1.3316	0.667	-1.996	0.046	-2.639	-0.024
C(WORK_GR)[T.3조]	-0.6299	0.669	-0.941	0.347	-1.942	0.682
C(WORK_GR)[T.4조]	-0.5604	0.620	-0.903	0.366	-1.776	0.656
PT_THK	-0.0635	0.044	-1.436	0.151	-0.150	0.023
PT_WDTH	-0.0007	0.001	-1.101	0.271	-0.002	0.001
PT_LTH	-6.362e-05	4.84e-05	-1.314	0.189	-0.000	3.12e-05
PT_WGT	1.614e-05	1.05e-05	1.540	0.123	-4.4e-06	3.67e-05
FUR_NO_ROW	0.0353	0.464	0.076	0.939	-0.874	0.945
FUR_HZ_TEMP	0.0541	0.029	1.840	0.066	-0.004	0.112
FUR_HZ_TIME	-0.0007	0.006	-0.103	0.918	-0.013	0.012
FUR_SZ_TEMP	-0.0735	0.057	-1.293	0.196	-0.185	0.038
FUR_SZ_TIME	-0.0187	0.012	-1.502	0.133	-0.043	0.006
FUR_TIME	-0.0019	0.007	-0.263	0.793	-0.016	0.012
ROLLING_TEMP_T5	0.0461	0.011	4.094	0.000	0.024	0.068
ROLLING_DESCALING	-0.7491	0.276	-2.712	0.007	-1.290	-0.208

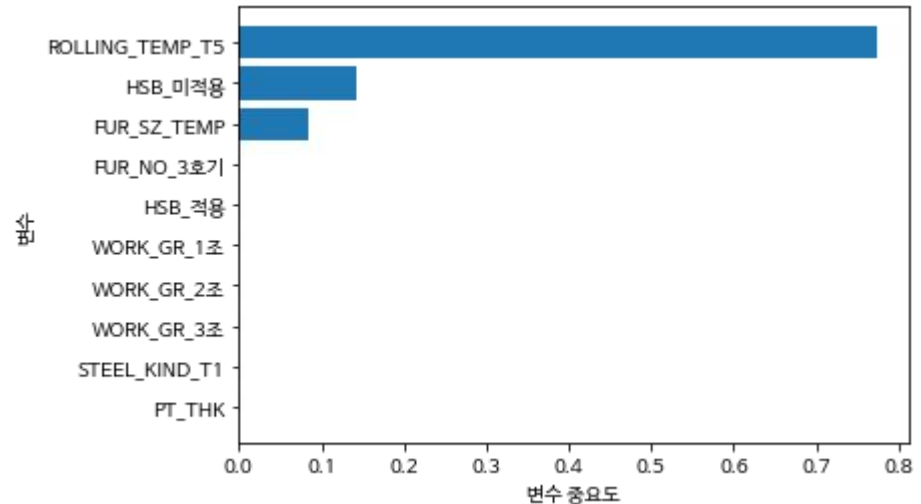
‘ROLLING_TEMP_T5’ 외의
 설명변수 의 p-value > 0.05

⇒ 로지스틱 회귀 분석으로
 변수의 중요도 평가가
 어려움

의사 결정 나무



Accuracy:0.965
Confusion matrix:
[[144 0]
[10 61]]



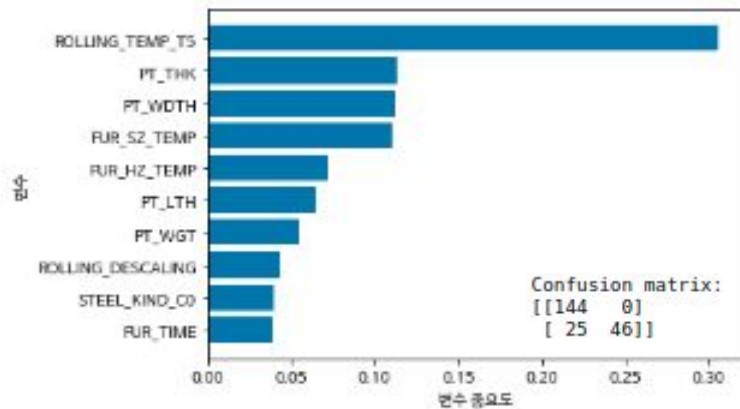
▲ 압연온도 > HSB 적용 여부, > 가열로 균열대 온도

◀ 모델의 정확도: 96.5%

◀ Confusion Matrix 결과, 불량률 양품으로 판정할 비율은 14%로 정확도 향상을 위해 파라미터의 변경이나 데이터의 추가가 필요

랜덤 포레스트/ 그라디언트 부스팅

● 랜덤 포레스트

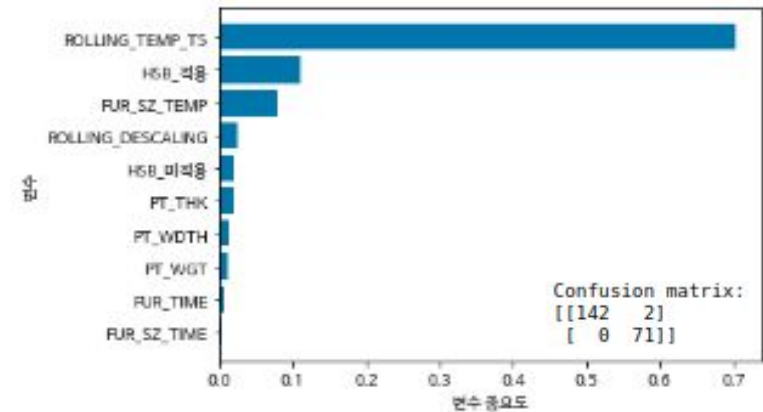


▲ 중요도 : 압연온도 > 판의 두께 > 판의 폭

◀ Train 정확도 : 91.2%,
Test 정확도 : 87.4%

◀ Confusion Matrix 결과, 불량인 제품을 양품으로 판정할 확률은 38%로 모델은 불량품을 분류하기에 부적절

● 그라디언트 부스팅

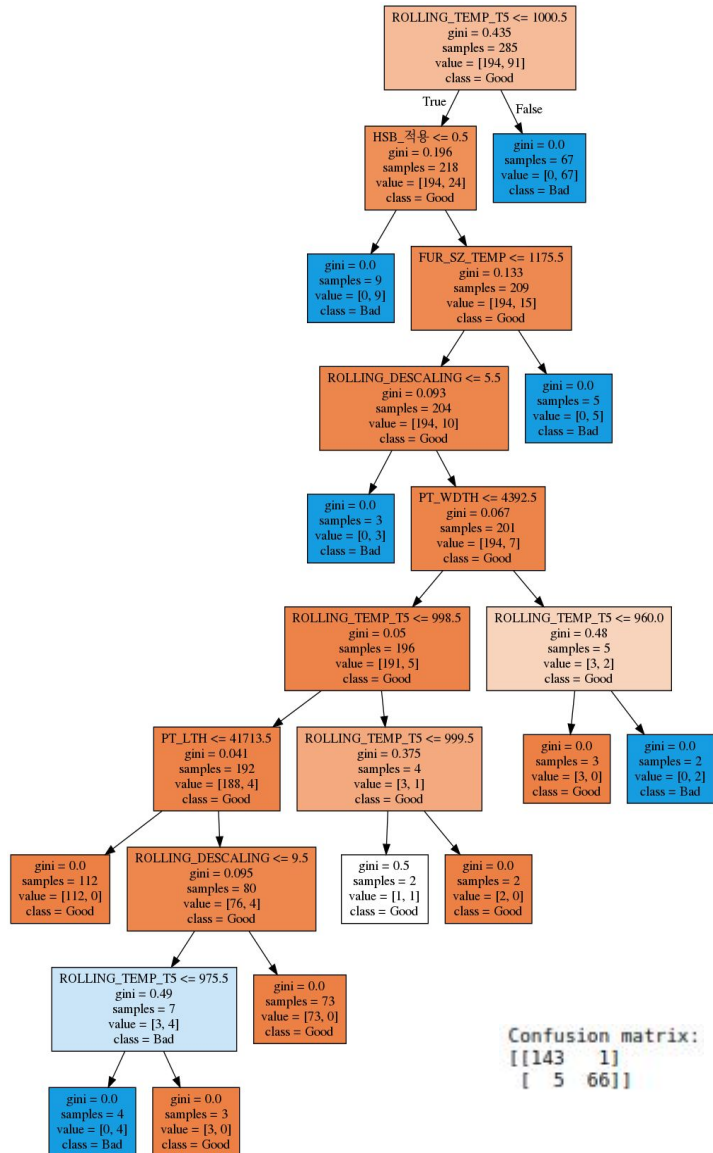


▲ 중요도 : 압연온도 > HSB적용 > 가열로 균열대 온도

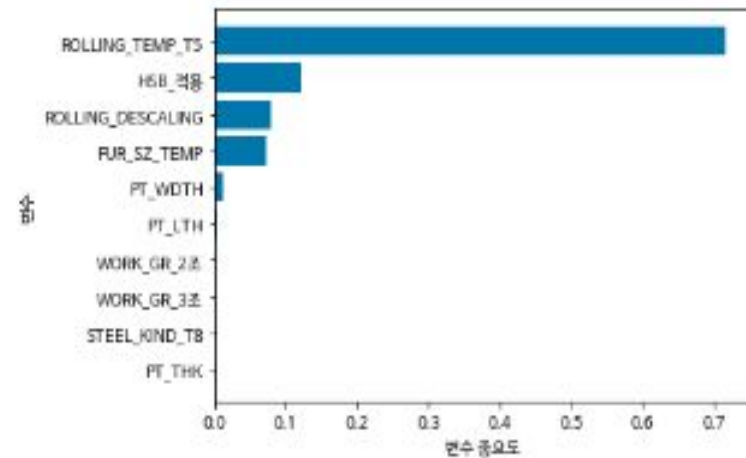
◀ Train 정확도 : 100%,
Test 정확도 : 99.1%

◀ Confusion Matrix 결과, 불량인 제품을 양품으로 판정할 확률은 0%로 모델은 불량품을 분류하기에 적절

중요도 높은 변수를 활용한 의사 결정 나무 모델



Confusion matrix:
[[143 1]
[5 66]]

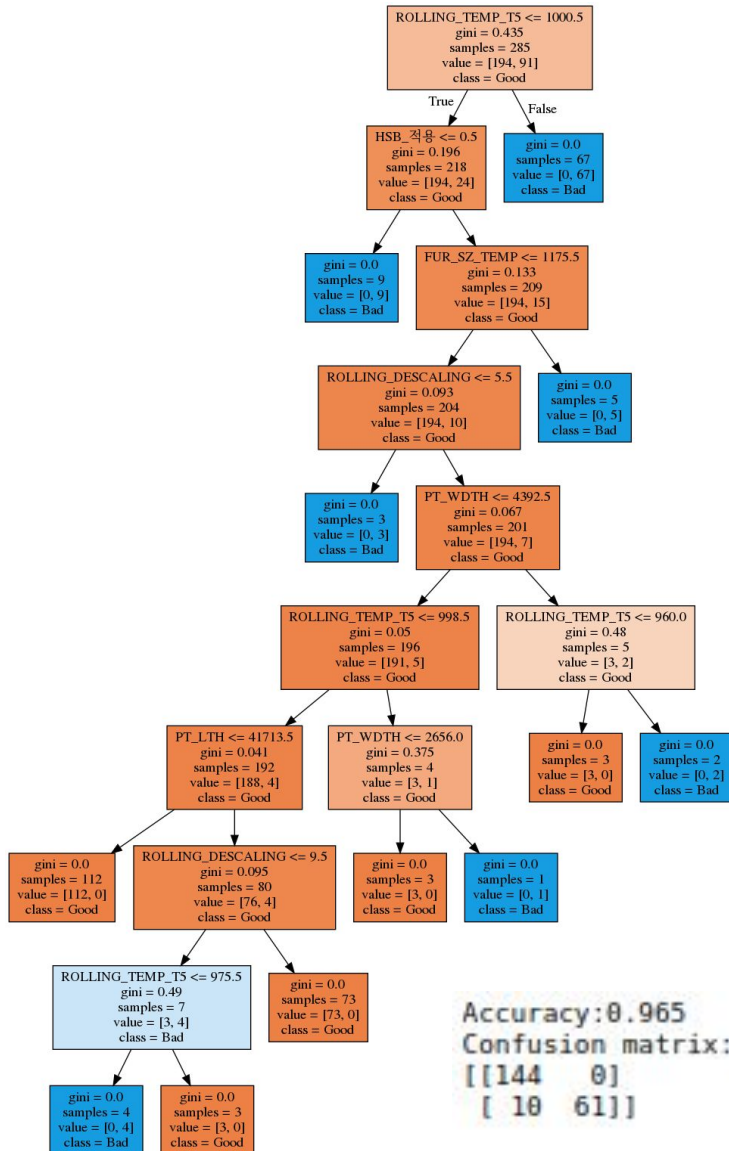


▲ 압연온도 > HSB 적용 여부 > 가열로 균열대 온도

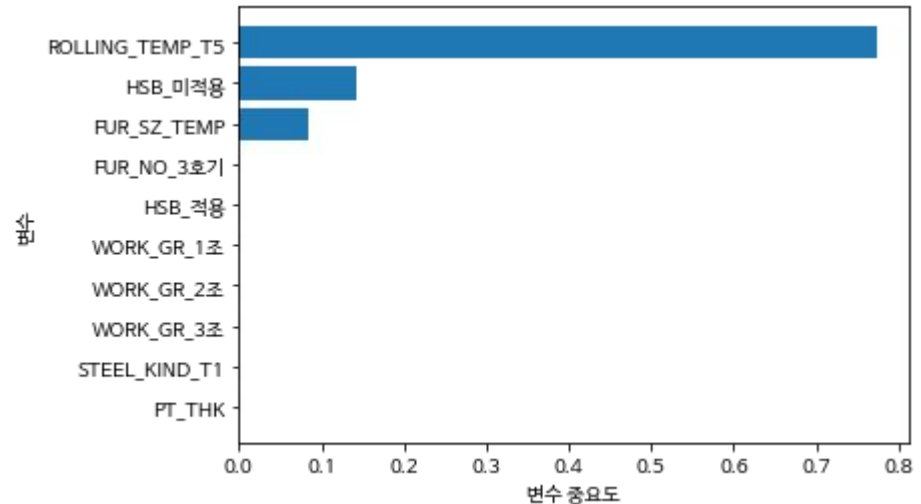
◀ 모델의 정확도: 99.6%

◀ Confusion Matrix 결과, 불량률 양품으로 판정할 비율은 7.04%로 정확도 향상을 위해 파라미터의 변경이나 데이터의 추가가 필요

의사 결정 나무



Accuracy:0.965
Confusion matrix:
[[144 0]
[10 61]]



▲ 압연온도 > HSB 적용 여부 > 압연 중 descaling 횟수

◀ 모델의 정확도: 96.5%

◀ Confusion Matrix 결과, 불량률 양품으로 판정할 비율은 14%로 정확도 향상을 위해 파라미터의 변경이나 데이터의 추가가 필요

분석 결과를 기준으로 분석 주제와 관련된 결론 또는 아이디어 제시 등을 제시합니다

- Plate의 물리적 특징은 회사 내에서 조절이 어렵지만 불량률에 중요한 요인을 미치므로 주문시 상대 회사에 권유 요구
- C0 강종의 경우 불량률이 높으므로 강종 종류의 변경이 요구 된다.
- HSB는 미적용시 불량률이 100%이므로 반드시 HSB 적용을 필요로 한다.
- 4번의 분류 모델에서 가장 중요한 변수는 압연 온도이다. 따라서 압연 온도는 불량률이 가장 낮은 800~900도로 유지해야 한다.
- 공정 과정 외에 작업조 차이에 따라 불량률의 차이를 보이므로 근로자에 대한 정확한 지침이 필요하다.

실습 과정을 통해 배운 또는 느낀 통찰, 아이디어, 애로사항 등을 정리합니다

- 후판에 대한 도메인 지식이 부족하여 문제를 이해하는 것에 어려움이 있었다.
- 스마트 팩토리 데이터를 직접 분석하면서 관련 과정에 대한 이해도가 깊어졌다.
- 회사 내에서 변경 가능한 변수에 대한 추가적인 데이터의 축척이 필요하다고 생각하였다.