스케일 발생 영향인자 분석 및 불량률 예측

과제 정의

- **압연공정이란?**
  - ➢ 금속 소성 가공 방법의 하나로 롤(roll)을 이용하여 두께를 줄이고 일정하게 만드는 과정

온도에 따른 분류

열간 압연

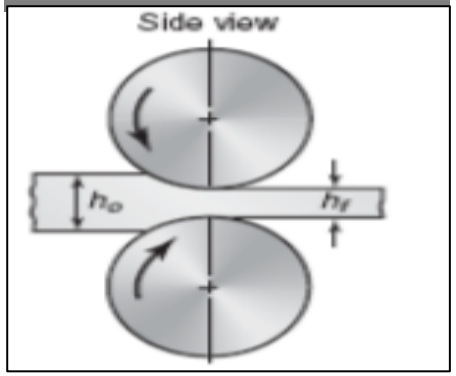
- 금속의 재결정 온도 이상에서 작업이 수행
- 재결정 이상에서 수행해서 가공경화가 일어나지 않는다.

냉간 압연

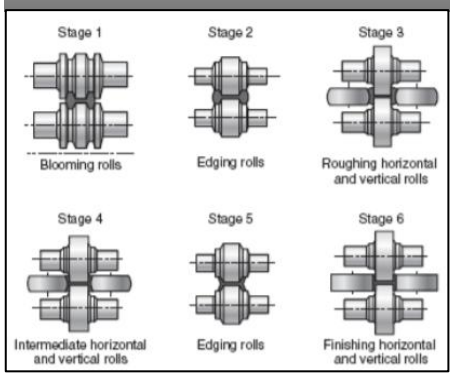
- 금속의 재결정 온도 이하에서 작업이 수행
- 가공경화에 의해 강도가 높아짐
- 열간압연에 비해서 두께를 많이 줄이지 못함

형상에 따른 분류

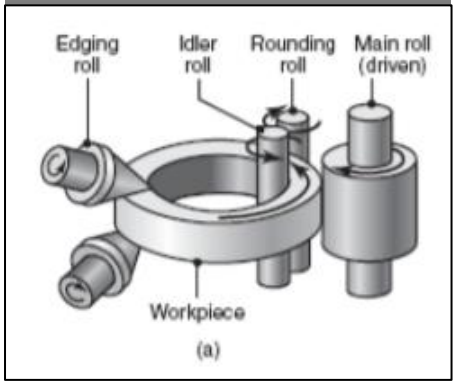
평판 압연



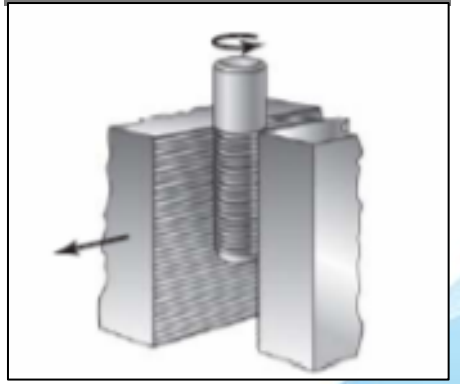
형상 압연



링 압연



나사 단조



과제 정의

- **스케일(Scale)이란?**

  - 물에 각종 광물질이나 가스 외에 다양한 이온(Ca++, Mg++, HCO3, SO4 등)이 존재

    -> 이들 이온의 화학적 결합물인 염류(CaCO3, MgCO3 등)가 온도상승 등으로 인한 용해도 감소로 침전

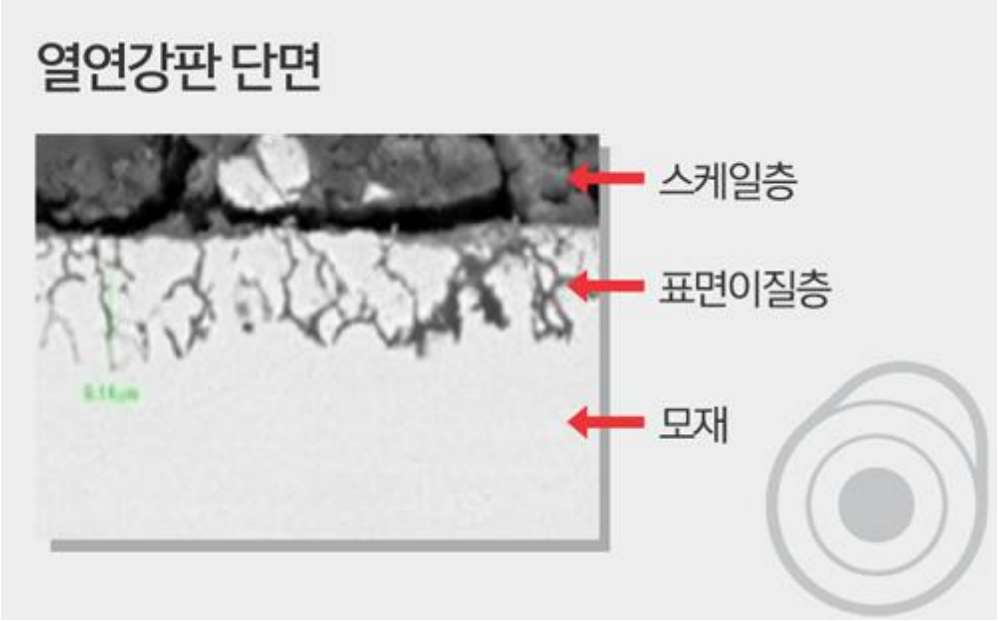
    -> 전열면의 표면이나 배관등에 부착

    -> 스케일(Scale) 생성

  - 스케일(Scale) 생성과정

    $Ca^{++} + 2(HCO_3^-) \rightarrow CaCO_3\,(\text{스케일}) + CO_2 + H_2O$

  - 압연과정에서 강판의 표면에 스케일층 형성
    -> 스케일(Scale) 불량 발생



과제 정의

• 분석 배경

- ✓ OO 공장의 고객사에서 최근 들어 'Scale 불량 발생 증가 ' 라는 이슈가 발생
- ✓ 원인 분석 결과 **압연공정**에서 Scale 불량이 급증
- ✓ 따라서 데이터 수집 후 다양한 분석 통해 **불량 발생의 근본 원인** 규명
- ✓ 결과를 해석하여 개선 기회를 도출

• 발생 현황

발생 원인	압입흠	Scratch	두께부족	Scale	계
발생률(%)	1.3	0.5	0.4	5.0	7.2%

• 잠재적인 인자 선정

PLATE_NO	Plate No	ID	범주형	Plate No	Nominal
ROLLING_DATE	작업시각	제외	연속형	작업시각	Datetime
SCALE	Scale불량	목표변수	범주형	Scale불량	Binary
SPEC	제품 규격	설명변수	범주형	제품 규격	Nominal
STEEL_KIND	강종	설명변수	범주형	강종	Nominal
PT_THICK	Plate 뚜께	설명변수	연속형	Plate 뚜께	Interval
PT_WIDTH	Plate 폭	설명변수	연속형	Plate 폭	Interval
PT_LENGTH	Plate 길이	설명변수	연속형	Plate 길이	Interval
PT_WEIGHT	Plate 중량	설명변수	연속형	Plate 중량	Interval
FUR_NO	가열로 호기	설명변수	범주형	가열로 호기	Nominal
FUR_NO_ROW	가열로 작업순번	설명변수	연속형	가열로	Interval
FUR_HZ_TEMP	가열로 가열대 온도	설명변수	연속형	가열로 가열대 온도	Interval
FUR_HZ_TIME	가열로 가열대 시간	설명변수	연속형	가열로 가열대 시간	Interval
FUR_SZ_TEMP	가열로 균열대 온도	설명변수	연속형	가열로 균열대 온도	Interval
FUR_SZ_TIME	가열로 균열대 시간	설명변수	연속형	가열로 균열대 시간	Interval
FUR_TIME	가열로 시간	설명변수	연속형	가열로 시간	Interval
FUR_EXTEMP	추출온도	설명변수	연속형	압연온도	Interval
ROLLING_TEMP_T5	압연온도	설명변수	연속형	가열대 온도	Interval
HSB	HSB적용(1-적용,0-미적용)	설명변수	범주형	HSB적용(1-적용,0-미적용)	Binary
ROLLING_DESCALING	압연 중 Descaling 횟수	설명변수	연속형	압연 중 Descaling 횟수	Interval
WORK_GR	작업조	설명변수	범주형	작업조	Nominal

With POSCO

분석 계획

- **데이터 현황**
  - 목표변수 : 스케일 불량 유무(SCALE)
  - 설명변수 : 제품 규격(SPEC), 강종(STEEL_KIND), Plate 두께(PT_THICK), Plate 폭(PT_WIDTH) 등 20개
  - 목표 변수는 범주형, 설명변수는 범주형 6개, 연속형 14개

PLATE_NO	ROLLING_DATE	SCALE	SPEC	STEEL_KIND	PT_THK	PT_WDTH	PT_LTH	PT_WGT	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
PB562774	2008-08-01:00:00:15	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	1	1144	116	1133	59	282	1133	934	적용	8	2조
PB562775	2008-08-01:00:00:16	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	2	1144	122	1135	53	283	1135	937	적용	8	2조
PB562776	2008-08-01:00:00:59	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	1	1129	116	1121	55	282	1121	889	적용	8	3조
PB562777	2008-08-01:00:01:24	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	2	1152	125	1127	68	316	1127	885	적용	8	3조
PB562778	2008-08-01:00:01:44	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	1	1140	134	1128	48	314	1128	873	적용	8	1조
PB562779	2008-08-01:00:02:06	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	2	1143	127	1128	57	314	1128	874	적용	8	4조
PB562780	2008-08-01:00:02:28	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	1호기	1	1138	126	1130	50	289	1130	878	적용	8	2조
PB562781	2008-08-01:00:02:21	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	1호기	2	1139	126	1131	52	294	1131	870	적용	8	4조
PB562782	2008-08-01:00:02:51	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	2호기	1	1127	126	1122	52	293	1122	873	적용	8	1조
PB562783	2008-08-01:00:03:15	양품	COMMON	T8	38.43	3129	16187	15280	2호기	2	1135	119	1124	73	298	1124	881	적용	8	4조
PB562784	2008-08-01:00:03:24	양품	COMMON	T8	38.43	3129	16187	15280	3호기	1	1127	134	1123	58	297	1123	869	적용	8	2조
PB562785	2008-08-01:00:04:15	불량	COMMON	T8	38.43	3129	16187	30560	3호기	2	1131	120	1125	68	299	1125	1057	적용	8	2조
PB562786	2008-08-01:00:04:20	양품	COMMON	T8	38.43	3129	16187	15280	1호기	1	1132	125	1127	62	290	1127	820	적용	8	3조
PB562787	2008-08-01:00:05:47	양품	COMMON	T0	30.23	1940	34797	16020	1호기	2	1119	130	1120	65	324	1120	926	적용	8	4조
PB562788	2008-08-01:00:05:25	양품	GL-E32-TM	T1	34.28	2207	30543	18140	2호기	1	1119	126	1119	72	311	1119	931	적용	8	3조
PB562789	2008-08-01:00:05:16	불량	GL-E32-TM	T1	50.46	2185	21767	37680	3호기	1	1127	127	1123	71	312	1123	929	적용	5	2조
PB562790	2008-08-01:01:10:14	양품	GL-E32-TM	T1	50.46	2200	21756	37920	2호기	2	1134	127	1124	92	329	1124	929	적용	6	2조
PB562791	2008-08-01:01:10:14	양품	GL-E32-TM	T1	50.46	2200	21756	37920	3호기	2	1124	117	1124	87	315	1124	929	적용	6	3조
PB562792	2008-08-01:01:10:44	양품	GL-E32-TM	T1	50.46	2200	21756	37920	1호기	1	1129	122	1125	78	313	1125	925	적용	6	2조
PB562793	2008-08-01:01:11:01	양품	GL-E32-TM	T1	50.46	2200	21756	37920	2호기	2	1124	54	1127	78	312	1127	928	적용	6	2조
PB562794	2008-08-01:01:11:08	양품	BV-EH36-TM	T8	50.46	2000	24500	38820	1호기	2	1110	123	1116	81	334	1116	860	적용	6	1조
PB562795	2008-08-01:01:12:45	양품	GL-E36-TM	T8	44.39	2040	27501	39100	2호기	1	1114	64	1120	82	335	1120	836	적용	6	4조
PB562796	2008-08-01:01:12:49	양품	GL-E36-TM	T8	44.39	2040	27501	39100	3호기	1	1113	124	1120	82	334	1120	832	적용	6	1조
PB562797	2008-08-01:01:13:47	양품	GL-E36-TM	T8	48.44	2095	24490	39020	3호기	2	1118	71	1124	86	336	1124	832	적용	6	3조
PB562798	2008-08-01:01:13:05	양품	GL-E36-TM	T8	48.44	2095	24490	39020	1호기	1	1117	69	1120	88	347	1120	832	적용	6	3조
PB562799	2008-08-01:01:14:20	양품	GL-E36-TM	T8	48.44	2000	25588	38920	1호기	2	1108	62	1117	97	351	1117	841	적용	6	3조
PB562800	2008-08-01:01:14:53	양품	COMMON	T1	45.4	2150	18453	14140	2호기	1	1123	62	1123	101	332	1123	933	적용	6	1조
PB562801	2008-08-01:01:14:25	양품	COMMON	T1	45.4	2150	18453	14140	2호기	2	1132	70	1126	95	344	1126	933	적용	6	1조
PB562802	2008-08-01:01:15:39	양품	COMMON	T1	45.4	2150	18453	14140	3호기	1	1129	43	1125	109	335	1125	937	적용	6	4조
PB562803	2008-08-01:01:15:14	양품	COMMON	T1	45.4	2090	18419	13720	3호기	2	1134	70	1127	94	335	1127	930	적용	6	1조
PB562804	2008-08-01:01:15:59	양품	COMMON	T1	45.4	2090	18419	13720	1호기	1	1124	78	1124	94	335	1124	936	적용	6	1조
PB562805	2008-08-01:01:15:34	양품	COMMON	T1	45.4	2090	18419	13720	1호기	2	1126	74	1125	92	334	1125	933	적용	6	1조
PB562806	2008-08-01:02:20:52	양품	COMMON	T8	44.9	3125	14008	15430	2호기	1	1111	61	1123	108	336	1123	838	적용	6	3조

- **탐색적 분석 계획**
  - 목표변수(범주형) - 설명변수(범주형) -> 히스토그램, 크로스탭, 카이제곱 검정
  - 목표변수(범주형) - 설명변수(연속형) -> 히스토그램
  - 설명변수(연속형) – 설명변수(연속형) -> 산점도, 상관분석

with POSCO

분석 계획

- **데이터 현황**
  - 목표변수 : 스케일 불량 유무(SCALE)
  - 설명변수 : 제품 규격(SPEC), 강종(STEEL_KIND), Plate 두께(PT_THICK), Plate 폭(PT_WIDTH) 등 20개
  - 목표 변수는 범주형, 설명변수는 범주형 6개, 연속형 14개

PLATE_NO	ROLLING_DATE	SCALE	SPEC	STEEL_KIND	PT_THK	PT_WDTH	PT_LTH	PT_WGT	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
PB562774	2008-08-01:00:00:15	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	1	1144	116	1133	59	282	1133	934	적용	8	2조
PB562775	2008-08-01:00:00:16	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	2	1144	122	1135	53	283	1135	937	적용	8	2조
PB562776	2008-08-01:00:00:59	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	1	1129	116	1121	55	282	1121	889	적용	8	3조
PB562777	2008-08-01:00:01:24	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	2	1152	125	1127	68	316	1127	885	적용	8	3조
PB562778	2008-08-01:00:01:44	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	1	1140	134	1128	48	314	1128	873	적용	8	1조
PB562779	2008-08-01:00:02:06	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	2	1143	127	1128	57	314	1128	874	적용	8	4조
PB562780	2008-08-01:00:02:28	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	1호기	1	1138	126	1130	50	289	1130	878	적용	8	2조
PB562781	2008-08-01:00:02:21	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	1호기	2	1139	126	1131	52	294	1131	870	적용	8	4조
PB562782	2008-08-01:00:02:51	양품	BV-EH36-TM	T8	38.33	3099	16719	15590	2호기	1	1127	126	1122	52	293	1122	873	적용	8	1조
PB562783	2008-08-01:00:03:15	양품	COMMON	T8	38.43	3129	16187	15280	2호기	2	1135	119	1124	73	298	1124	881	적용	8	4조
PB562784	2008-08-01:00:03:24	양품	COMMON	T8	38.43	3129	16187	15280	3호기	1	1127	134	1123	58	297	1123	869	적용	8	2조
PB562785	2008-08-01:00:04:15	불량	COMMON	T8	38.43	3129	16187	30560	3호기	2	1131	120	1125	68	299	1125	1057	적용	8	2조
PB562786	2008-08-01:00:04:20	양품	COMMON	T8	38.43	3129	16187	15280	1호기	1	1132	125	1127	62	290	1127	820	적용	8	3조
PB562787	2008-08-01:00:05:47	양품	COMMON	T0	30.23	1940	34797	16020	1호기	2	1119	130	1120	65	324	1120	926	적용	8	4조
PB562788	2008-08-01:00:05:25	양품	GL-E32-TM	T1	34.28	2207	30543	18140	2호기	1	1119	126	1119	72	311	1119	931	적용	8	3조
PB562789	2008-08-01:00:05:16	불량	GL-E32-TM	T1	50.46	2185	21767	37680	3호기	1	1127	127	1123	71	312	1123	929	적용	5	2조
PB562790	2008-08-01:01:10:14	양품	GL-E32-TM	T1	50.46	2200	21756	37920	2호기	2	1134	127	1124	92	329	1124	929	적용	6	2조
PB562791	2008-08-01:01:10:14	양품	GL-E32-TM	T1	50.46	2200	21756	37920	3호기	2	1124	117	1124	87	315	1124	929	적용	6	3조
PB562792	2008-08-01:01:10:44	양품	GL-E32-TM	T1	50.46	2200	21756	37920	1호기	1	1129	122	1125	78	313	1125	925	적용	6	2조
PB562793	2008-08-01:01:11:01	양품	GL-E32-TM	T1	50.46	2200	21756	37920	2호기	2	1124	54	1127	78	312	1127	928	적용	6	2조
PB562794	2008-08-01:01:11:08	양품	BV-EH36-TM	T8	50.46	2000	24500	38820	1호기	2	1110	123	1116	81	334	1116	860	적용	6	1조
PB562795	2008-08-01:01:12:45	양품	GL-E36-TM	T8	44.39	2040	27501	39100	2호기	1	1114	64	1120	82	335	1120	836	적용	6	4조
PB562796	2008-08-01:01:12:49	양품	GL-E36-TM	T8	44.39	2040	27501	39100	3호기	1	1113	124	1120	82	334	1120	832	적용	6	1조
PB562797	2008-08-01:01:13:47	양품	GL-E36-TM	T8	48.44	2095	24490	39020	3호기	2	1118	71	1124	86	336	1124	832	적용	6	3조
PB562798	2008-08-01:01:13:05	양품	GL-E36-TM	T8	48.44	2095	24490	39020	1호기	1	1117	69	1120	88	347	1120	832	적용	6	3조
PB562799	2008-08-01:01:14:20	양품	GL-E36-TM	T8	48.44	2000	25588	38920	1호기	2	1108	62	1117	97	351	1117	841	적용	6	3조
PB562800	2008-08-01:01:14:53	양품	COMMON	T1	45.4	2150	18453	14140	2호기	1	1123	62	1123	101	332	1123	933	적용	6	1조
PB562801	2008-08-01:01:14:25	양품	COMMON	T1	45.4	2150	18453	14140	2호기	2	1132	70	1126	95	344	1126	933	적용	6	1조
PB562802	2008-08-01:01:15:39	양품	COMMON	T1	45.4	2150	18453	14140	3호기	1	1129	43	1125	109	335	1125	937	적용	6	4조
PB562803	2008-08-01:01:15:14	양품	COMMON	T1	45.4	2090	18419	13720	3호기	2	1134	70	1127	94	335	1127	930	적용	6	1조
PB562804	2008-08-01:01:15:59	양품	COMMON	T1	45.4	2090	18419	13720	1호기	1	1124	78	1124	94	335	1124	936	적용	6	1조
PB562805	2008-08-01:01:15:34	양품	COMMON	T1	45.4	2090	18419	13720	1호기	2	1126	74	1125	92	334	1125	933	적용	6	1조
PB562806	2008-08-01:02:20:52	양품	COMMON	T8	44.9	3125	14008	15430	2호기	1	1111	61	1123	108	336	1123	838	적용	6	3조

- **탐색적 분석 계획**
  - 목표변수(범주형) - 설명변수(범주형) -> 히스토그램, 크로스탭, 카이제곱 검정
  - 목표변수(범주형) - 설명변수(연속형) -> 히스토그램
  - 설명변수(연속형) – 설명변수(연속형) -> 산점도, 상관분석

with POSCO

분석 계획

• 모델링 분석 계획

- ➢ 로지스틱 회귀, 의사결정나무, 랜덤포레스트, 그래디언트 부스팅 기법을 사용하여 목표변수를 예측하는 모델 생성
- ➢ 생성된 모델들의 train/test accuracy, F1 score, AUC 분석
  -> 모델들이 잘 생성되었는지 평가하고 서로 비교
- ➢ 생성된 모델들의 설명변수 중요도 확인
  -> 설명변수들이 목표변수에 미치는 영향력 확인

• 중요인자 도출

- ➢ 탐색적 분석과 모델링 분석을 통해 설명변수들의 중요도 측정
- ➢ 가장 영향력이 높은 설명변수들을 선택

• 결론 및 대안 제시

- ➢ 중요변수들을 가설과 비교
- ➢ 중요변수들을 토대로 SCALE 불량을 막을 수 있는 방법을 고안

With POSCO

데이터 정제

- **필요없는 변수 제거**
  - ➢ PLATE_NO는 스케일 불량 유무(SCALE) 와 관련이 없으므로 데이터에서 제외
  - ➢ 작업시각(ROLLING_DATE)은 스케일 불량 유무(SCALE)와 관련이 없으므로 데이터에서 제외

```
# 'PLATE_NO'와 'ROLLING_DATE'는 목표변수와 관련없는 변수이므로 제거
df_raw.drop("PLATE_NO", axis = 1, inplace = True)
df_raw.drop("ROLLING_DATE", axis = 1, inplace = True)
df_raw
```

	SCALE	SPEC	STEEL_KIND	PT_THK	PT_WDTH	PT_LTH	PT_WGT	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
0	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	1	1144	116	1133	59	282	1133	934	적용	8	2조
1	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	2	1144	122	1135	53	283	1135	937	적용	8	2조
2	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	1	1129	116	1121	55	282	1121	889	적용	8	3조
3	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	2	1152	125	1127	68	316	1127	885	적용	8	3조
4	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	1	1140	134	1128	48	314	1128	873	적용	8	1조
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
715	불량	NK-KA	C0	20.14	3580	38639	21870	3호기	1	1172	72	1164	62	245	1164	1005	적용	8	2조
716	양품	NV-A32	C0	15.08	3212	48233	18340	2호기	1	1150	61	1169	61	238	1169	947	적용	10	1조
717	양품	NV-A32	C0	16.60	3441	43688	19590	2호기	2	1169	65	1163	77	247	1163	948	적용	10	4조
718	양품	LR-A	C0	15.59	3363	48740	80240	3호기	2	1179	86	1163	45	243	1163	940	적용	10	2조
719	양품	GL-A32	C0	16.09	3400	54209	69840	3호기	1	1186	82	1169	45	239	1169	957	적용	10	2조

➢ 'PLATE_NO', 'ROLLING_DATE' 가 제외 -> 19개 변수 존재 (목표변수 : 1개, 설명변수 : 18개)

with POSCO

데이터 정제

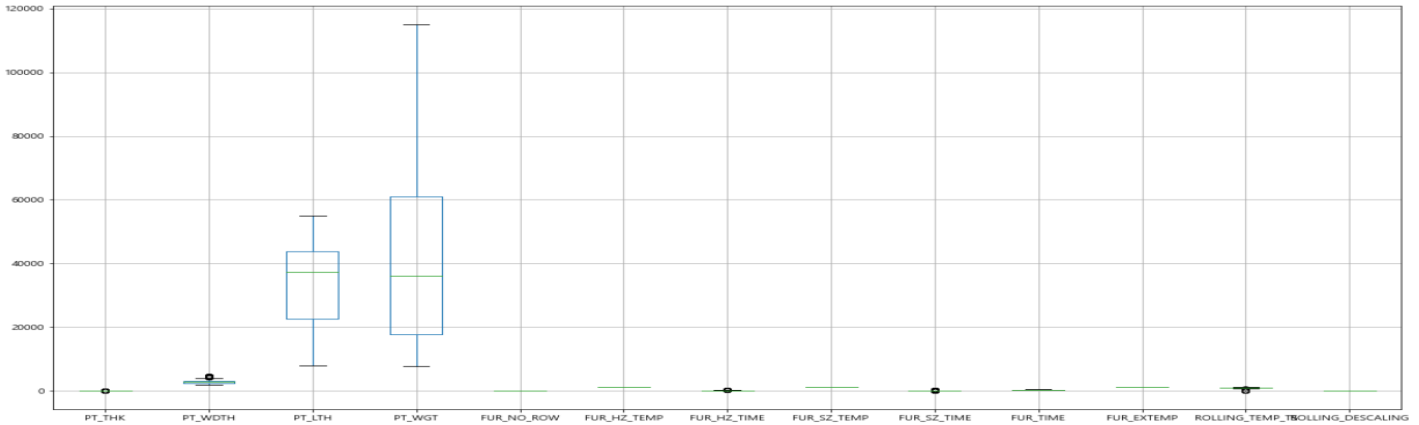
- **결측치 확인**

  ➢ 결측치가 보이지 않음

```
# 결측치 확인
df_raw.isnull().sum()
```

```
SCALE              0
SPEC               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
```

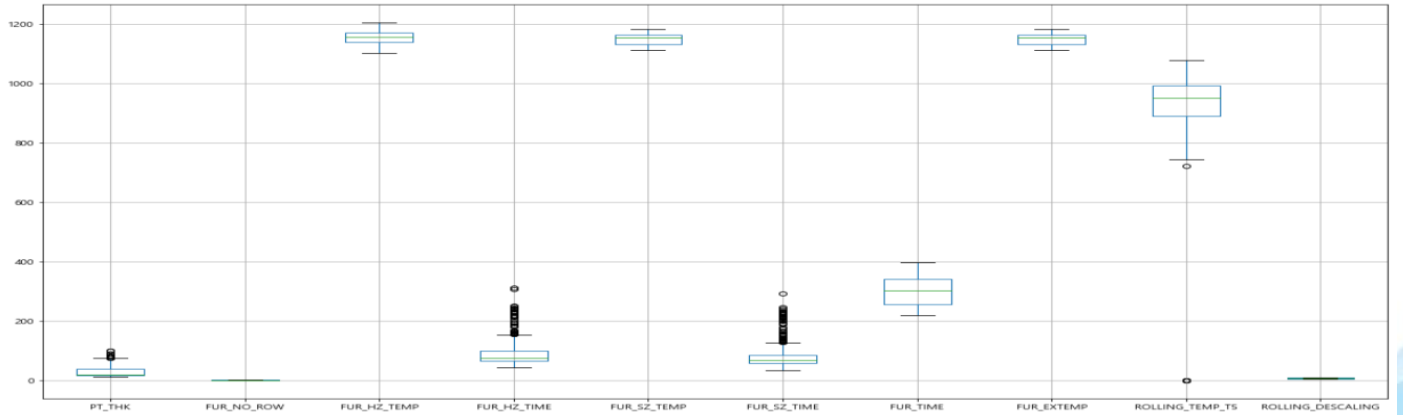
- **이상치 확인**



➢ 연속형 변수들의 Boxplot 생성 결과 몇몇 변수들의 boxplot 크기 때문에 다른 변수들의 boxplot 결과가 보이지 않음

➢ 크기가 큰 boxplot을 생성하는 ' PT_LTH', ' PT_WGT', ' PT_WDTH' 은 이상치가 보이지 않음

-> ' PT_LTH', ' PT_WGT', 'PT_WDTH' 제외 후 나머지 변수들의 boxplot 생성

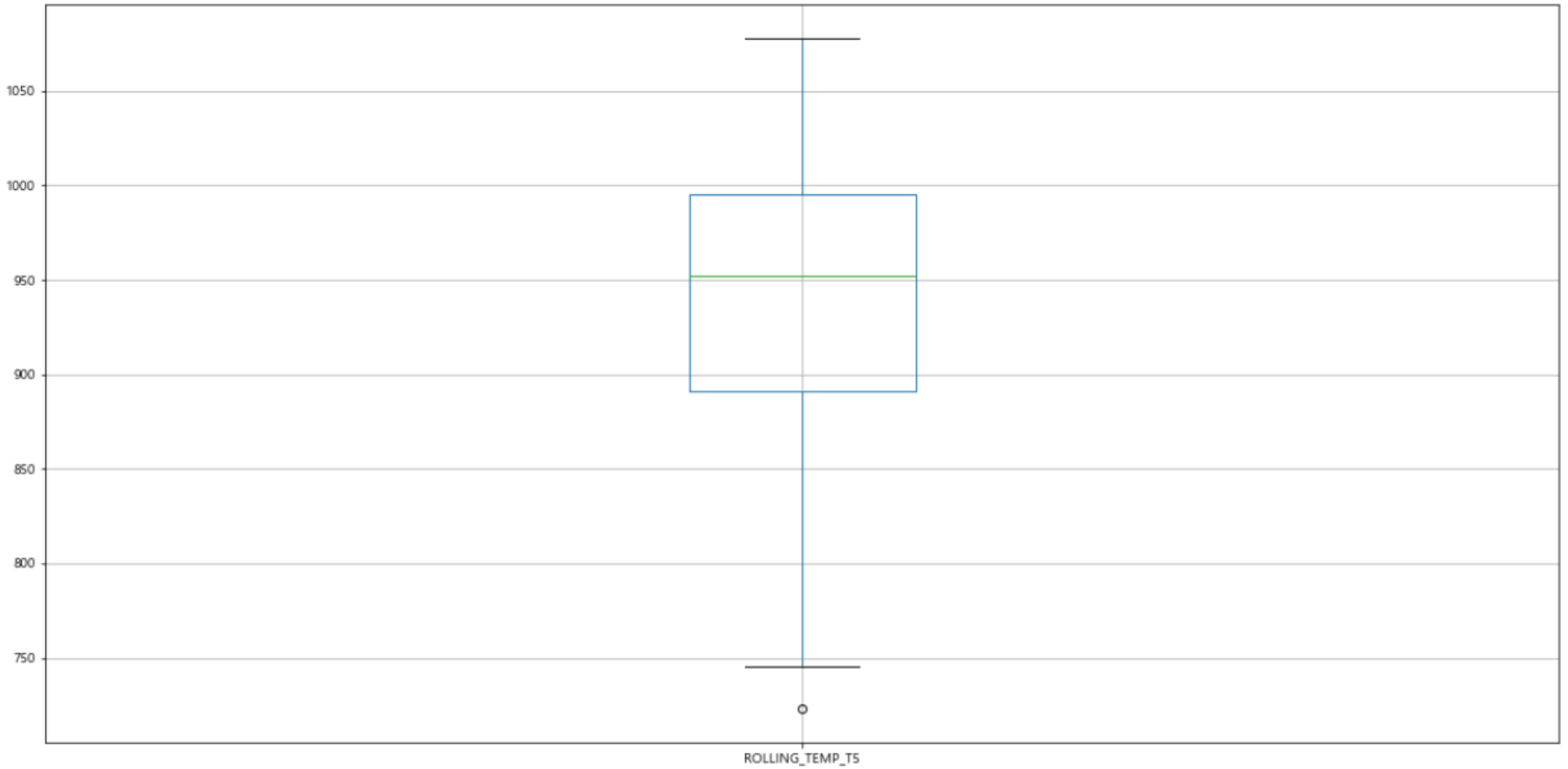


➢ 확인 결과 압연온도(ROLLING_TEMP_T5)에서 이상치가 발견

with POSCO

데이터 정제

- **이상치 제거**
  - 압연온도(ROLLING_TEMP_T5)가 100 이하인 값들을 제거
  - Boxplot을 그려본 결과 이상치가 제거됨

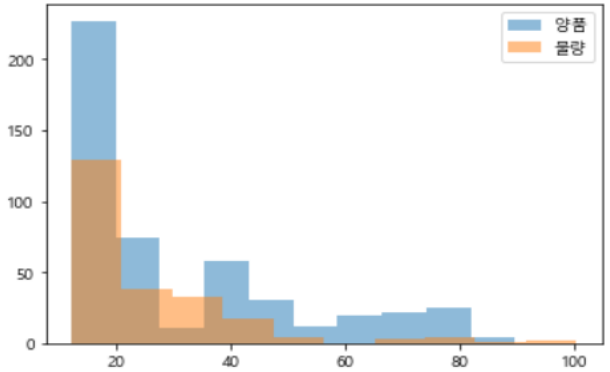


With POSCO

탐색적 분석

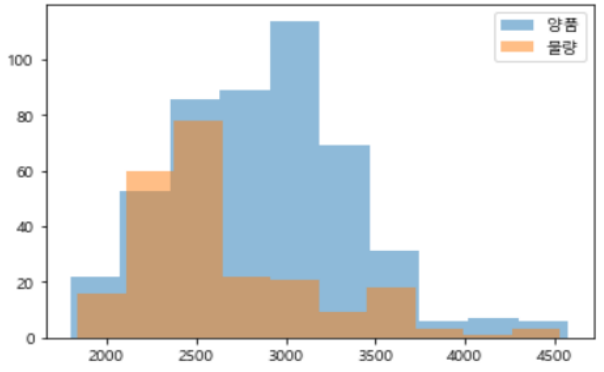
- **그래프를 통한 분석**
  - ➢ 목표변수가 범주형 데이터 -> 목표변수와 설명변수의 히스토그램 생성 후 분석

1. Scale과 PT_THK의 관계



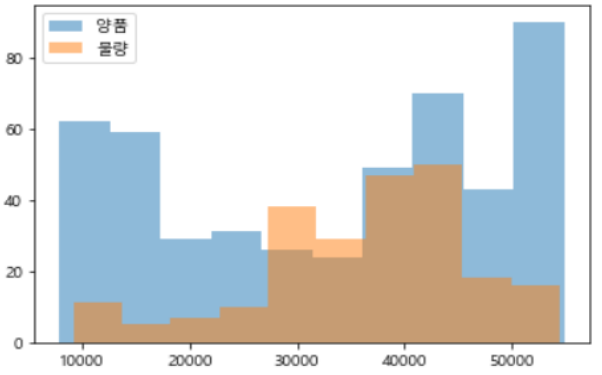
➢ Plate 두께(PT_THK)가 얇아질수록 Scale 불량률 증가

2. Scale과 PT_WDTH의 관계



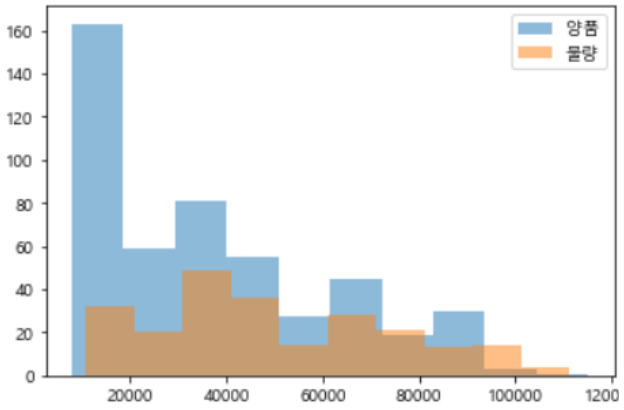
➢ Plate 폭(PT_WDTH)이 얇아질수록 SCALE 불량률 증가

3. Scale과 PT_WDTH의 관계



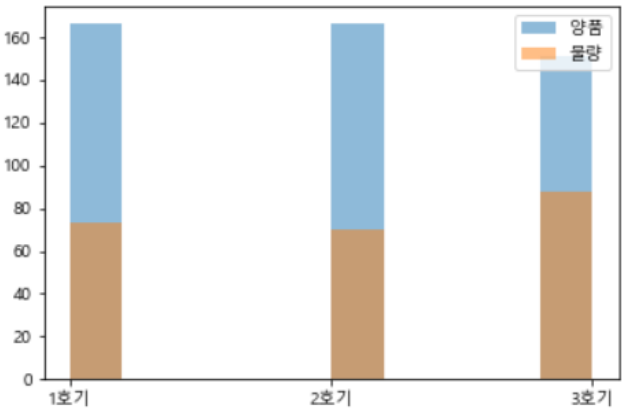
➢ Plate 길이(PT_LTH)가 30000~45000정도 일 때 SCALE 불량률 증가

4. Scale과 PT_WGT의 관계



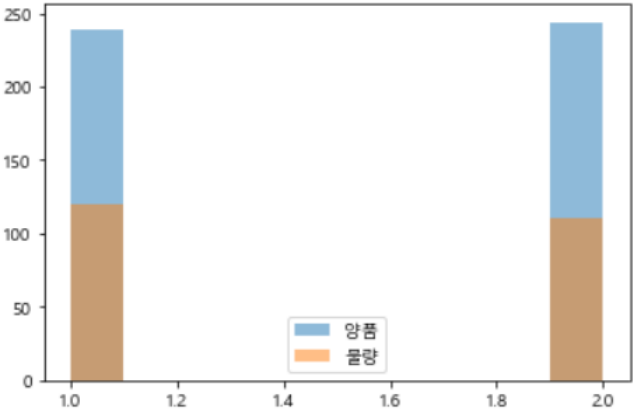
➢ Plate 중량(PT_WGT)이 늘어날수록 SCALE 불량률 증가

5. Scale과 FUR_NO의 관계



➢ 가열로 호기(FUR_NO) 중 3호기에서 SCALE 불량률이 높음

6. Scale과 FUR_NO_ROW의 관계



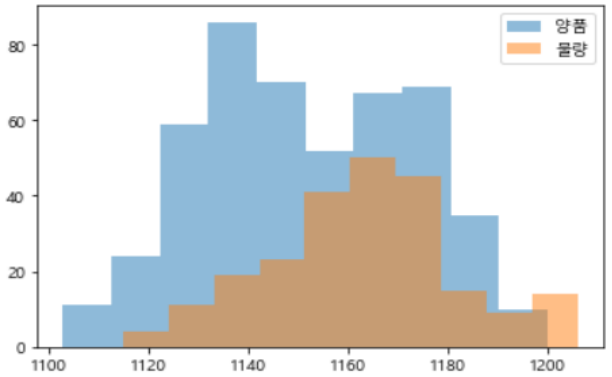
➢ 가열로 작업 순번(FUR_NO_ROW)은 SCALE 불량률에 영향을 미치지 않음

with POSCO

탐색적 분석

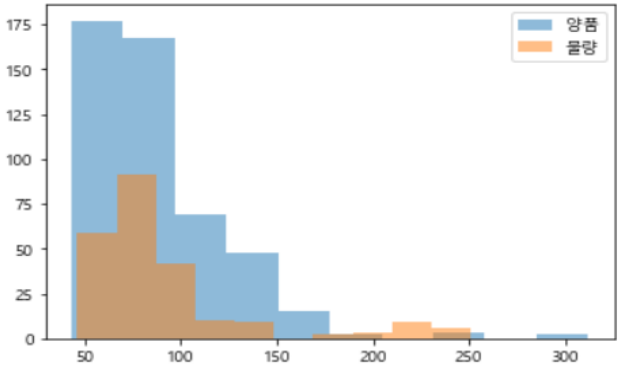
- **그래프를 통한 분석**

7. Scale과 FUR_HZ_TEMP의 관계



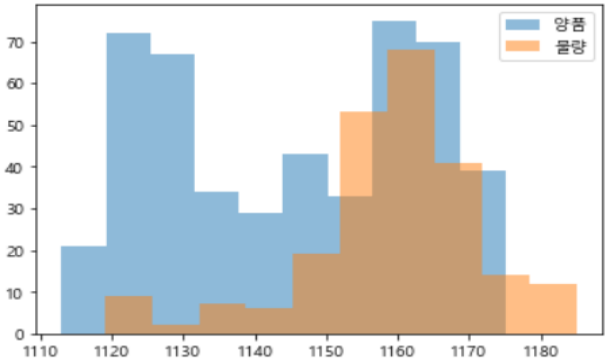
➢ 가열로 가열대 온도(FUR_HZ_TEMP)가 1160~1180 사이일 때 SCALE 불량률 증가

8. Scale과 FUR_HZ_TIME의 관계



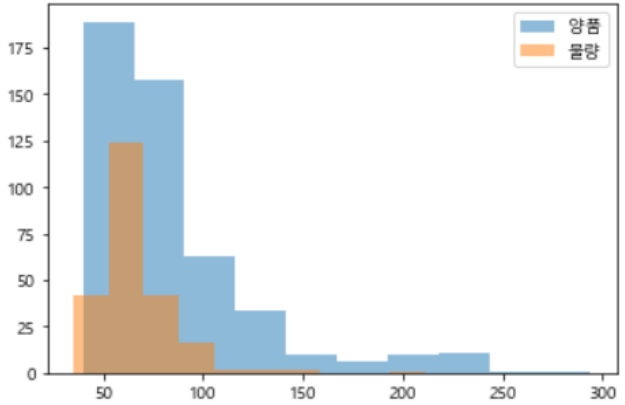
➢ 가열로 가열대 시간(FUR_HZ_TIME)과 SCALE 불량률은 상관성을 보이지 않음

9. Scale과 FUR_SZ_TEMP의 관계



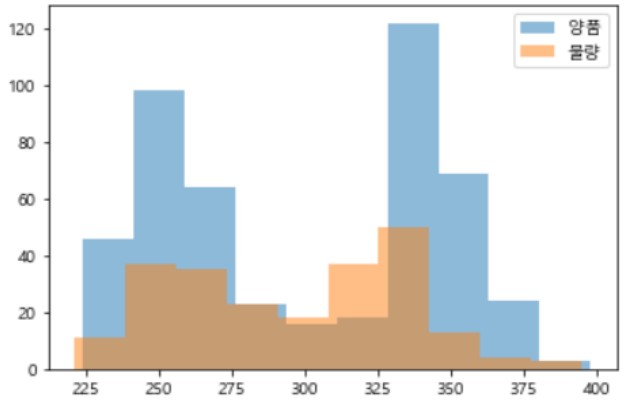
➢ 가열로 균열대 온도(FUR_SZ_TEMP)가 높을수록 SCALE 불량률 증가

10. Scale과 PT_WGT의 관계



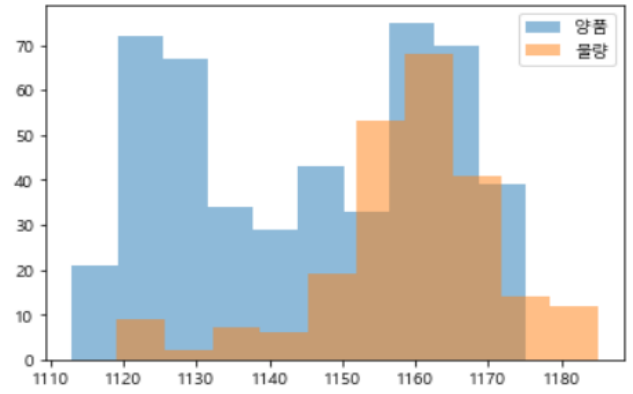
➢ 가열로 균열대 시간(FUR_SZ_TIME)이 낮을수록 SCALE 불량률 증가

11. Scale과 FUR_TIME의 관계



➢ 가열로 시간(FUR_TIME)이 275~325일 때 SCALE 불량률이 증가

12. Scale과 FUR_EXTEMP의 관계



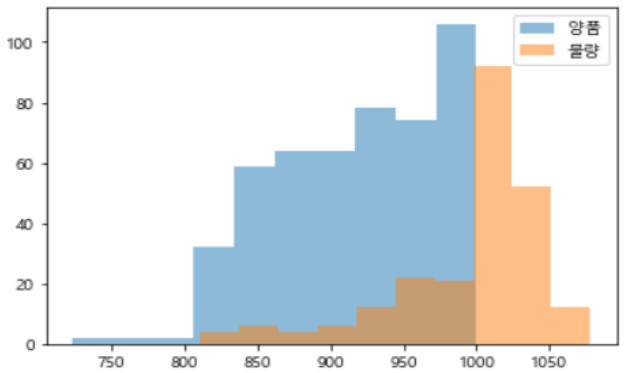
➢ 추출온도(FUR_EXTEMP)가 증가할수록 SCALE 불량률 증가

with POSCO

탐색적 분석

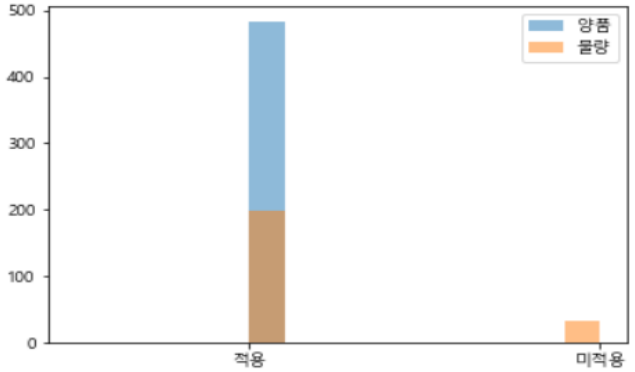
- **검정을 통한 분석**

13. Scale과 ROLLING_TEMP_T5 관계



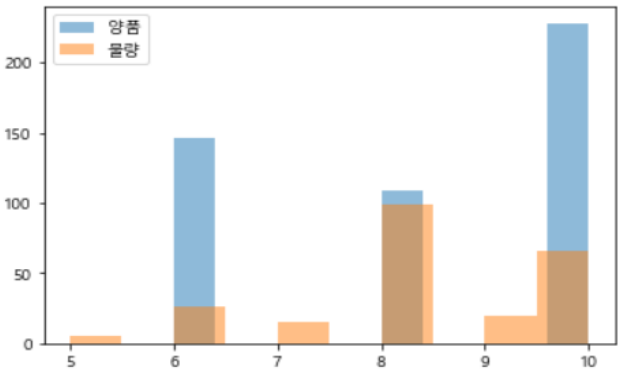
- 압연온도(ROLLING_TEMP_T5)가 높을수록 SCALE 불량률 증가

14. Scale과 HSB의 관계



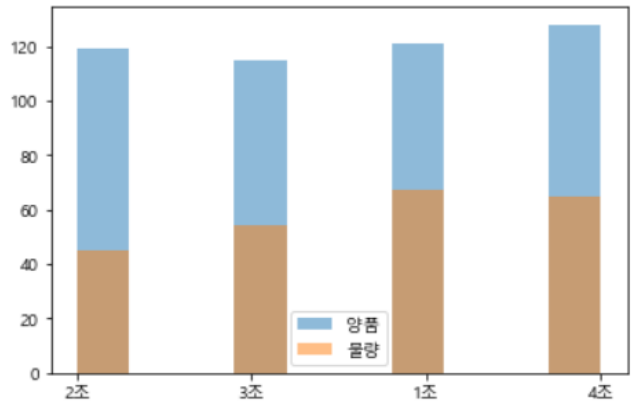
- HSB를 미적용하면 SCALE불량이 100% 발생
- 따라서 SCALE을 줄이기 위해선 HSB를 필수 적용

15. Scale과 ROLLING_DESCALING의 관계



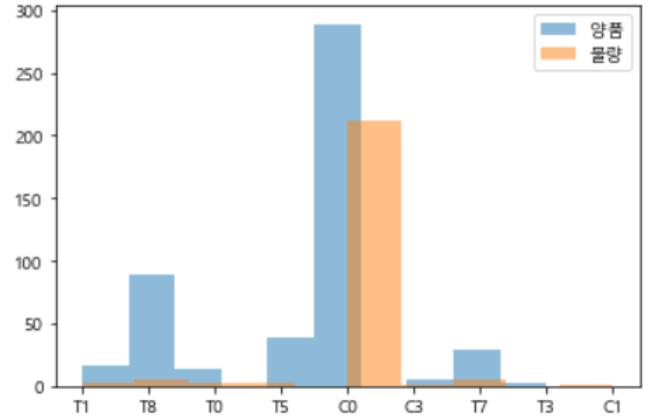
- 압연 중 Descaling 횟수(ROLLING_DESCALING)가 8일때 SCALE 불량률이 가장 높음

16. Scale과 WORK_GR의 관계



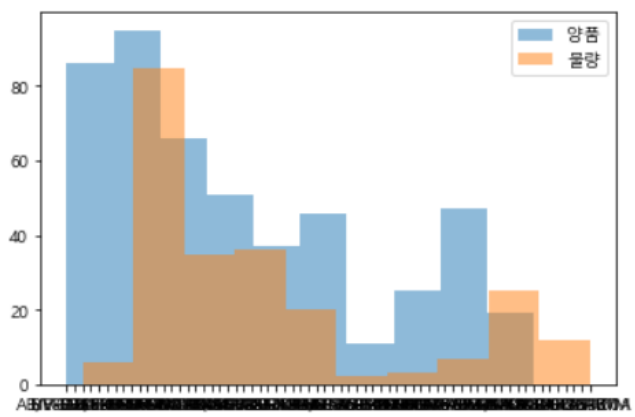
- 작업조(WORK_GR) 중 1조의 겨우 SCALE 불량률이 가장 높음

17. Scale과 FUR_TIME의 관계



- 강송(STEEL_KIND)이 C0일 때 물량률이 가장 높음

18. Scale과 SPEC의 관계



- 각 제품규격(SPEC)마다 SCALE 불량률이 다름

with POSCO

- **검정을 통한 분석**

1) 교차표

➢ 목표변수(범주형)과 설명변수(범주형)간의 관계 분석

1. STEEL_KIND와 SCALE의 관계

STEEL_KIND	C0	C1	C3	T0	T1	T3	T5	T7	T8
SCALE									
불량	212	1	1	2	2	0	2	6	5
양품	289	0	6	13	16	2	39	29	89

STEEL_KIND	C0	C1	C3	T0	T1	T3	T5	₩
SCALE								
불량	0.423154	1.0	0.142857	0.133333	0.111111	0.0	0.04878	
양품	0.576846	0.0	0.857143	0.866667	0.888889	1.0	0.95122	

STEEL_KIND	T7	T8
SCALE		
불량	0.171429	0.053191
양품	0.828571	0.946809

✓ C1은 SCALE불량일 확률이 100%, 그러나 1개만 존재.

✓ 따라서 C1의 불량률은 신용할 수 없음

✓ 따라서 C1을 제외하고 불량률이 가장 높은 것은 CO

2. FUR_NO와 SCALE의 관계

FUR_NO	1호기	2호기	3호기
SCALE			
불량	73	70	88
양품	166	166	151

FUR_NO	1호기	2호기	3호기
SCALE			
불량	0.305439	0.29661	0.368201
양품	0.694561	0.70339	0.631799

✓ 3호기에서 SCALE 불량이 가장 많이 발생하는 것을 알 수 있다.

with POSCO

- **검정을 통한 분석**

1) 교차표

➢ 목표변수(범주형)과 설명변수(범주형)간의 관계 분석

3. HSB와 SCALE의 관계

HSB SCALE	미적용	적용
불량	33	198
양품	0	483

HSB SCALE	미적용	적용
불량	1.0	0.290749
양품	0.0	0.709251

✓ HSB가 미적용되었을 때 불량일 확률 = 1

✓ SCALE불량을 일으키지 않기 위해서는 HSB를 필수 적용할 필요 있음

4. WORK_GR와 SCALE의 관계

WORK_GR SCALE	1조	2조	3조	4조
불량	67	45	54	65
양품	121	119	115	128

WORK_GR SCALE	1조	2조	3조	4조
불량	0.356383	0.27439	0.319527	0.336788
양품	0.643617	0.72561	0.680473	0.663212

✓ 1조에서 SCALE불량률이 가장 많이 나옴

✓ 2조에서는 SCALE불량률이 가장 적게 나옴

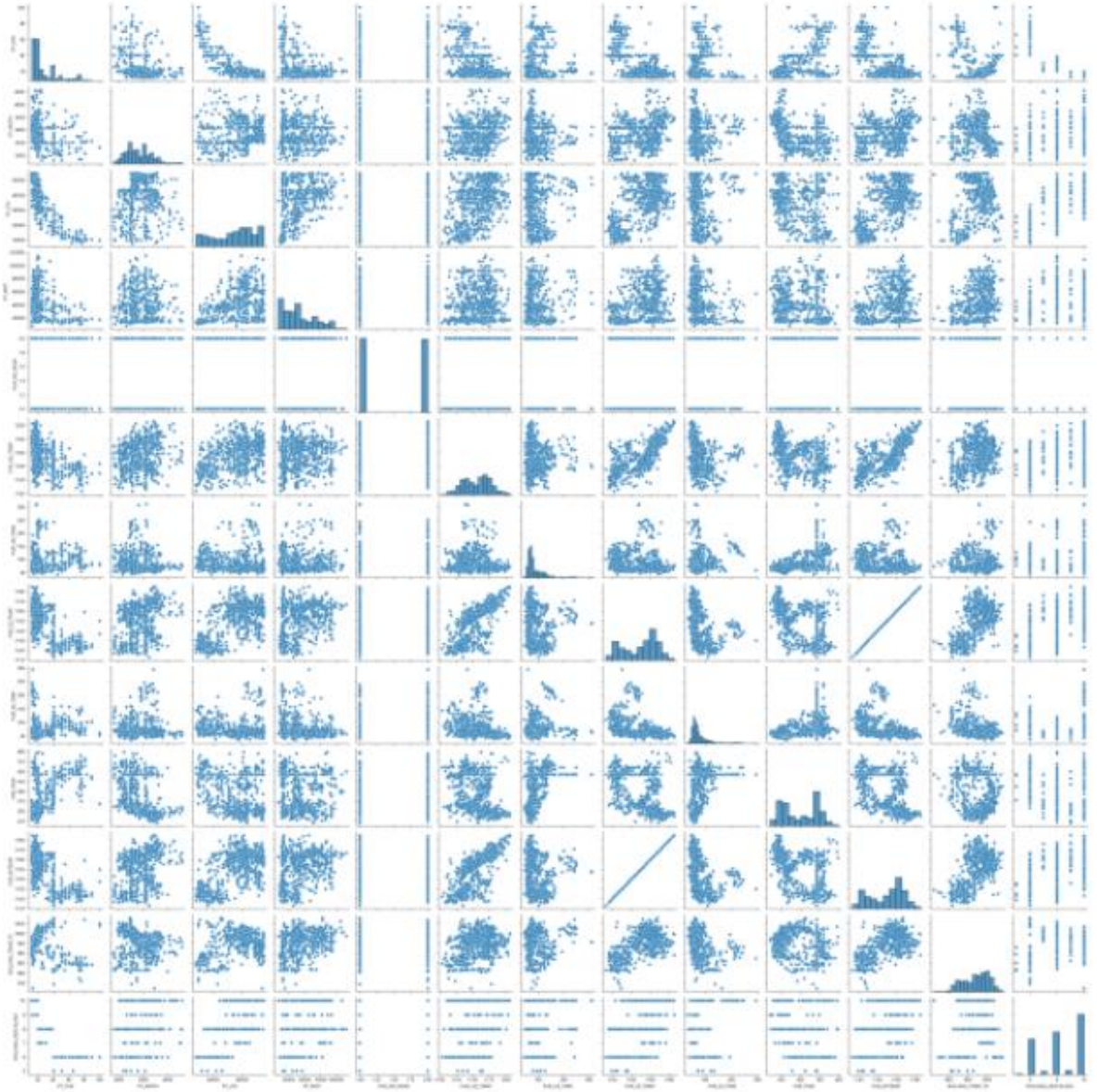
with POSCO

탐색적 분석

- **검정을 통한 분석**

2) 산점도 행렬

- 연속형 설명 변수들간의 산점도
- Plate 길이(PT_LTH)와 Plate 두께(PT_THK)가 서로 음의 상관관계를 보이는 등 몇몇 설명변수들간에 상관관계가 존재
- FUR_SZ_TEMP(가열로 균열대 온도)와 FUR_EXTEMP(압연온도)가 완벽히 비례
  → 둘 중 하나만 사용해도 될 것 같음



with POSCO

- 검정을 통한 분석

3) 상관관계 표

	PT_THK	PT_WDTH	PT_LTH	PT_WGT	FUR_NO_ROW	FUR_HZ_TEMP	FUR_HZ_TIME	FUR_SZ_TEMP	FUR_SZ_TIME	FUR_TIME	FUR_EXTEMP	ROLLING_TEMP_T5	ROLLING_DESCALING
PT_THK	1.000	-0.314	-0.862	-0.394	-0.004	-0.520	0.160	-0.692	0.118	0.406	-0.692	-0.503	-0.836
PT_WDTH	-0.314	1.000	0.121	0.030	0.003	0.181	-0.122	0.229	0.019	-0.393	0.229	-0.113	0.343
PT_LTH	-0.862	0.121	1.000	0.449	-0.020	0.468	-0.075	0.641	-0.072	-0.245	0.641	0.434	0.807
PT_WGT	-0.394	0.030	0.449	1.000	-0.022	0.150	0.012	0.356	-0.192	-0.054	0.356	0.420	0.234
FUR_NO_ROW	-0.004	0.003	-0.020	-0.022	1.000	0.005	-0.015	0.008	0.049	0.018	0.008	-0.007	0.018
FUR_HZ_TEMP	-0.520	0.181	0.468	0.150	0.005	1.000	-0.112	0.770	-0.225	-0.342	0.770	0.356	0.465
FUR_HZ_TIME	0.160	-0.122	-0.075	0.012	-0.015	-0.112	1.000	-0.201	0.178	0.475	-0.201	0.006	-0.135
FUR_SZ_TEMP	-0.692	0.229	0.641	0.356	0.008	0.770	-0.201	1.000	-0.456	-0.471	1.000	0.662	0.644
FUR_SZ_TIME	0.118	0.019	-0.072	-0.192	0.049	-0.225	0.178	-0.456	1.000	0.449	-0.456	-0.379	-0.107
FUR_TIME	0.406	-0.393	-0.245	-0.054	0.018	-0.342	0.475	-0.471	0.449	1.000	-0.471	-0.210	-0.360
FUR_EXTEMP	-0.692	0.229	0.641	0.356	0.008	0.770	-0.201	1.000	-0.456	-0.471	1.000	0.662	0.644
ROLLING_TEMP_T5	-0.503	-0.113	0.434	0.420	-0.007	0.356	0.006	0.662	-0.379	-0.210	0.662	1.000	0.370
ROLLING_DESCALING	-0.836	0.343	0.807	0.234	0.018	0.465	-0.135	0.644	-0.107	-0.360	0.644	0.370	1.000

- 연속형 설명 변수들간의 상관관계 분석
- Plate 길이(PT_LTH)와 Plate 두께(PT_THK)의 상관관계가 -0.862로 음의 상관관계를 보이는 등 산점도 행렬에서 살펴본 변수들간의 선형관계를 수치로 확인
- FUR_SZ_TEMP(가열로 균열대 온도)와 FUR_EXTEMP(압연온도)가 완벽히 비례
  -> 둘 중 하나를 제거하기로 결정

- **검정을 통한 분석**

※ 설명변수 FUR_EXTEMP 제거

➢ FUR_SZ_TEMP(가열로 균열대 온도)와 FUR_EXTEMP(압연온도)가 겹치므로 FUR_EXTEMP를 제거

	SCALE	SPEC	STEEL_KIND	PT_THK	PT_WDTH	PT_LTH	PT_WGT	FUR_NO	FUR_NO_ROW	FUR_HZ_TEMP	FUR_HZ_TIME	FUR_SZ_TEMP	FUR_SZ_TIME	FUR_TIME	ROLLING_TEMP_T5	HSB	ROLLING_DESCALING	WORK_GR
0	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	1	1144	116	1133	59	282	934	적용	8	2조
1	양품	AB/EH32-TM	T1	32.25	3707	15109	14180	1호기	2	1144	122	1135	53	283	937	적용	8	2조
2	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	1	1129	116	1121	55	282	889	적용	8	3조
3	양품	NV-E36-TM	T8	33.27	3619	19181	18130	2호기	2	1152	125	1127	68	316	885	적용	8	3조
4	양품	BV-EH36-TM	T8	38.33	3098	13334	12430	3호기	1	1140	134	1128	48	314	873	적용	8	1조

➢ FUR_EXTEMP 제거 후 설명변수(범주형) : 6개, 목표변수(연속형) : 11개

4) 카이제곱 독립성 검정

➢ 카이 제곱 독립 검정은 보통 두 범주형(명목형) 변수 사이에 유의미한 관계가 있는지 아닌지를 검정

➢ 범주형 변수인 SPEC과 STEEL_KIND의 경우 값이 너무 다양하므로 카이제곱 검정을 통해 목표변수 SCALE과 상관성이 없는 범주값들을 제거

➢ 카이제곱 검정의 귀무 가설(H0), 대립 가설(H1)

귀무 가설(H0)	**두 변수들 간에는 관계가 없다.**
대립 가설(H1)	**두 변수들 간에는 관계가 있다.**

- 검정을 통한 분석

4) 카이제곱 독립성 검정

a. SPEC과 SCALE의 카이제곱 검정

➢ SPEC의 더미 변수들 생성

	SCALE	SPEC_A131-DH36TM	SPEC_A283-C	SPEC_A516-60	SPEC_A709-36	SPEC_AB/A	SPEC_AB/AH32	SPEC_AB/B	SPEC_AB/EH32-TM	SPEC_AB/EH36-TM	...
0	1	0	0	0	0	0	0	0	1	0	...
1	1	0	0	0	0	0	0	0	1	0	...
2	1	0	0	0	0	0	0	0	0	0	...
3	1	0	0	0	0	0	0	0	0	0	...
4	1	0	0	0	0	0	0	0	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...
715	0	0	0	0	0	0	0	0	0	0	...
716	1	0	0	0	0	0	0	0	0	0	...
717	1	0	0	0	0	0	0	0	0	0	...
718	1	0	0	0	0	0	0	0	0	0	...
719	1	0	0	0	0	0	0	0	0	0	...

➢ 카이제곱 검정 결과 p-value가 0.05 이상인 범주값 컬럼들을 제거

➢ P-value가 0.05이하인 범주값들 ->

```
['SPEC_A283-C',
 'SPEC_AB/EH36-TM',
 'SPEC_BV-EH36-TM',
 'SPEC_COMMON',
 'SPEC_GL-A36-TM',
 'SPEC_JS-SM490A',
 'SPEC_JS-SM490YB',
 'SPEC_JS-SS400',
 'SPEC_KR-A',
 'SPEC_KS-SM490A',
 'SPEC_PILAC-BT33']
```

➢ 카이제곱 검정 후 데이터 현황

(오른쪽 아래 표)

...	SPEC_AB/EH36-TM	SPEC_BV-EH36-TM	SPEC_COMMON	SPEC_GL-A36-TM	SPEC_JS-SM490A	SPEC_JS-SM490YB	SPEC_JS-SS400	SPEC_KR-A	SPEC_KS-SM490A	SPEC_PILAC-BT33
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	1	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0

with POSCO

- 검정을 통한 분석

4) 카이제곱 독립성 검정

a. STEEL_KIND와 SCALE의 카이제곱 검정

➢ STEEL_KIND의 더미 변수들 생성

	SCALE	STEEL_KIND_C0	STEEL_KIND_C1	STEEL_KIND_C3	STEEL_KIND_T0	STEEL_KIND_T1	STEEL_KIND_T3	STEEL_KIND_T5	STEEL_KIND_T7	STEEL_KIND_T8
0	1	0	0	0	0	1	0	0	0	0
1	1	0	0	0	0	1	0	0	0	0
2	1	0	0	0	0	0	0	0	0	1
3	1	0	0	0	0	0	0	0	0	1
4	1	0	0	0	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...
715	0	1	0	0	0	0	0	0	0	0
716	1	1	0	0	0	0	0	0	0	0
717	1	1	0	0	0	0	0	0	0	0
718	1	1	0	0	0	0	0	0	0	0
719	1	1	0	0	0	0	0	0	0	0

➢ 카이제곱 검정 결과 p-value가 0.05 이상인 범주값 컬럼들을 제거

➢ P-value가 0.05이하인 범주값들

```
['STEEL_KIND_C0', 'STEEL_KIND_T5', 'STEEL_KIND_T8']
```

➢ 카이제곱 검정 후 데이터 현황

(오른쪽 아래 표)

...	SPEC_GL-A36-TM	SPEC_JS-SM490A	SPEC_JS-SM490YB	SPEC_JS-SS400	SPEC_KR-A	SPEC_KS-SM490A	SPEC_PILAC-BT33	STEEL_KIND_C0	STEEL_KIND_T5	STEEL_KIND_T8
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	0
...	0	0	0	0	0	0	0	0	0	1
...	0	0	0	0	0	0	0	0	0	1
...	0	0	0	0	0	0	0	0	0	1
...	...	...	...	...	...	...	...	...	...	...
...	0	0	0	0	0	0	0	1	0	0
...	0	0	0	0	0	0	0	1	0	0
...	0	0	0	0	0	0	0	1	0	0
...	0	0	0	0	0	0	0	1	0	0
...	0	0	0	0	0	0	0	1	0	0

탐색적 분석

- **검정을 통한 분석**

4) 카이제곱 독립성 검정

c. 카이제곱 검정 후 데이터 현황

```
Index(['SCALE', 'PT_THK', 'PT_WDTH', 'PT_LTH', 'PT_WGT', 'FUR_NO',
       'FUR_NO_ROW', 'FUR_HZ_TEMP', 'FUR_HZ_TIME', 'FUR_SZ_TEMP',
       'FUR_SZ_TIME', 'FUR_TIME', 'ROLLING_TEMP_T5', 'HSB',
       'ROLLING_DESCALING', 'WORK_GR', 'SPEC_A283-C', 'SPEC_AB/EH36-TM',
       'SPEC_BV-EH36-TM', 'SPEC_COMMON', 'SPEC_GL-A36-TM', 'SPEC_JS-SM490A',
       'SPEC_JS-SM490YB', 'SPEC_JS-SS400', 'SPEC_KR-A', 'SPEC_KS-SM490A',
       'SPEC_PILAC-BT33', 'STEEL_KIND_CO', 'STEEL_KIND_T5', 'STEEL_KIND_T8'],
      dtype='object')
```

➢ 원래 설명변수들에서 SPEC, STEEL_KIND가 빠지고 SPEC, STEEL_KIND 내에 있는 범주 값들 중 일부가 one-hot 벡터 형태로 추가된 것을 알 수 있다.

- **검정 후 중요인자 선정**

➢ 가열로 작업순번, 가열로 가열대 시간을 제외한 나머지 설명변수들은 목표변수와 상관성이 존재하므로 잠재인자가 이들 중 있다고 판단

with POSCO

1. 로지스틱 회귀 분석

1) Train/test set 설정

➢ Train/test set을 7:3 비율로 생성

```
# 데이터들을 train/test data로 분리
df_train, df_test = train_test_split(df_raw, test_size = 0.3, random_state = 1234)

print('train data size : {}'.format(df_train.shape))
print('test data size : {}'.format(df_test.shape))
```

```
train data size : (499, 30)
test data size : (215, 30)
```

2) 선형회귀 모델 생성

➢ 모델의 설명력 : 0.5534

➢ p-value가 0.05 이하인 변수

-> WORK_GR(C(WORK_GR)[T.2조], C(WORK_GR)[T.3조]), PT_THK, FUR_HZ_TEMP, FUR_SZ_TEMP, ROLLING_TEMP_T5, ROLLING_DESCALING

➢ WORK_GR은 유의하다고 판단할 수 있음

➢ Train/test accuracy와 Confusion Matrix

```
Train Accuracy:0.881764

Test Accuracy:0.823256

Confusion Matrix:
[[125  19]
 [ 19  52]]
```

```
Current function value: 0.280145
Iterations: 35
Function evaluations: 50
Gradient evaluations: 39
                           Logit Regression Results
==============================================================================
Dep. Variable:                  SCALE   No. Observations:                  499
Model:                          Logit   Df Residuals:                      473
Method:                           MLE   Df Model:                           25
Date:                Tue, 09 Mar 2021   Pseudo R-squ.:                  0.5534
Time:                        22:17:50   Log-Likelihood:                -139.79
converged:                      False   LL-Null:                       -313.05
Covariance Type:            nonrobust   LLR p-value:                2.478e-58
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.4685     26.010     -0.018      0.986     -51.447      50.510
C(FUR_NO)[T.2호기]   -0.2539      0.392     -0.648      0.517      -1.022       0.515
C(FUR_NO)[T.3호기]    0.4529      0.399      1.135      0.256      -0.329       1.235
C(HSB)[T.적용]    -14.3019     18.249     -0.784      0.433     -50.069      21.465
C(WORK_GR)[T.2조]   -1.5667      0.462     -3.388      0.001      -2.473      -0.660
C(WORK_GR)[T.3조]   -1.7638      0.509     -3.466      0.001      -2.761      -0.766
C(WORK_GR)[T.4조]   -0.6136      0.420     -1.460      0.144      -1.437       0.210
PT_THK        -0.0622      0.024     -2.619      0.009      -0.109      -0.016
PT_WDTH       -0.0006      0.000     -1.337      0.181      -0.001       0.000
PT_LTH     -4.673e-05   2.93e-05     -1.596      0.110      -0.000    1.06e-05
PT_WGT     1.092e-07   7.88e-06      0.014      0.989   -1.53e-05    1.56e-05
FUR_NO_ROW    -0.0930      0.314     -0.296      0.767      -0.709       0.523
FUR_HZ_TEMP    0.0492      0.018      2.808      0.005       0.015       0.084
FUR_HZ_TIME    0.0048      0.005      0.991      0.322      -0.005       0.014
FUR_SZ_TEMP   -0.0634      0.031     -2.033      0.042      -0.125      -0.002
FUR_TIME      -0.0037      0.005     -0.819      0.413      -0.013       0.005
ROLLING_TEMP_T5    0.0431      0.006      6.666      0.000       0.030       0.056
ROLLING_DESCALING   -0.5861      0.193     -3.044      0.002      -0.963      -0.209
SPEC_A283_C    0.5202      1.444      0.360      0.719      -2.310       3.350
SPEC_ABEH36_TM    0.5609      2.810      0.200      0.842      -4.947       6.068
SPEC_JS_SM490A    1.4951      1.208      1.237      0.216      -0.873       3.864
SPEC_JS_SM490YB   -0.3050      0.536     -0.569      0.569      -1.356       0.746
SPEC_JS_SS400    0.8460      1.297      0.652      0.514      -1.697       3.389
SPEC_KR_A     -0.6016      0.668     -0.900      0.368      -1.912       0.708
SPEC_KS_SM490A    0.4961      2.288      0.217      0.828      -3.988       4.980
SPEC_PILAC_BT33   -2.1747      1.365     -1.593      0.111      -4.850       0.500
==============================================================================
```


with POSCO

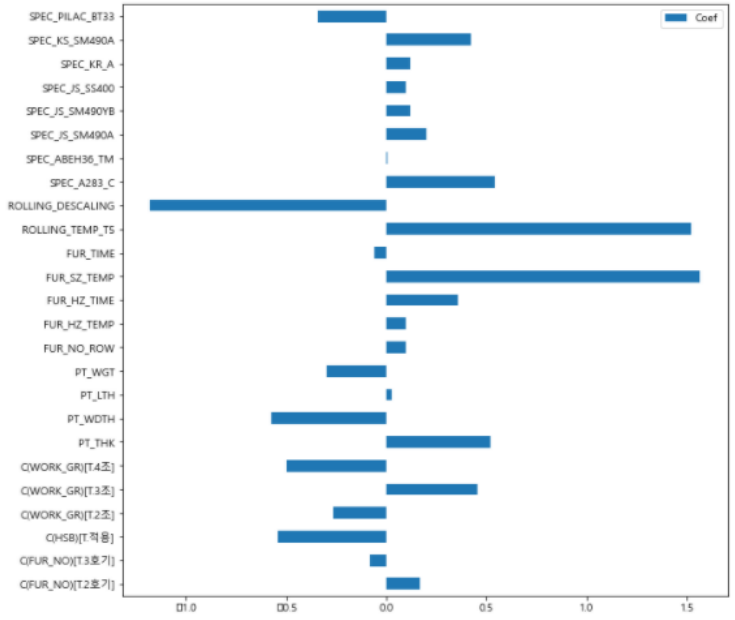
1. 로지스틱 회귀 분석

3) 데이터 표준화(Scaling)

- 연속형 설명변수들을 표준화
- 표준화한 목표변수(SCALE)에서 값이 0 이상이면 1, 아닌 경우 0으로 설정

SCALE	PT_THK	PT_WDTH	PT_LTH	PT_WGT	FUR_NO_ROW	FUR_HZ_TEMP	FUR_HZ_TIME	FUR_SZ_TEMP	FUR_SZ_TIME	...
0	0.165524	1.738825	-1.399249	-1.102663	-0.994413	-0.550140	0.726729	-0.958762	-0.544799	...
0	0.165524	1.738825	-1.399249	-1.102663	1.005618	-0.550140	0.885652	-0.844450	-0.709067	...
0	0.219261	1.568260	-1.103953	-0.943272	-0.994413	-1.270256	0.726729	-1.644633	-0.654311	...
0	0.219261	1.568260	-1.103953	-0.943272	1.005618	-0.166077	0.965114	-1.301698	-0.298397	...
0	0.485835	0.558436	-1.527970	-1.173279	-0.994413	-0.742171	1.203499	-1.244542	-0.845957	...

4) 설명변수 중요도 확인



- FUR_HZ_TIME,
  ROLLING_TEMP_T5,
  ROLLING_DESCALIING,
  PT_WIDTH,
  HSB(적용)
  순으로 영향력이 강함

With POSCO

2. 의사결정나무

1) Train/test set 설정

- 설명변수와 목표변수를 나눔
- Train/test set을 7:3 비율로 생성

```
# 데이터들을 train/test data로 분리
df_train_x, df_test_x, df_train_y, df_test_y = train_test_split(df_raw_x, df_raw_y,
                                                                test_size = 0.3, random_state = 1234)
print('train data X size : {}'.format(df_train_x.shape))
print('train data Y size : {}'.format(df_train_y.shape))
print('test data X size : {}'.format(df_test_x.shape))
print('test data Y size : {}'.format(df_test_y.shape))
```

```
train data X size : (499, 35)
train data Y size : (499, 1)
test data X size : (215, 35)
test data Y size : (215, 1)
```

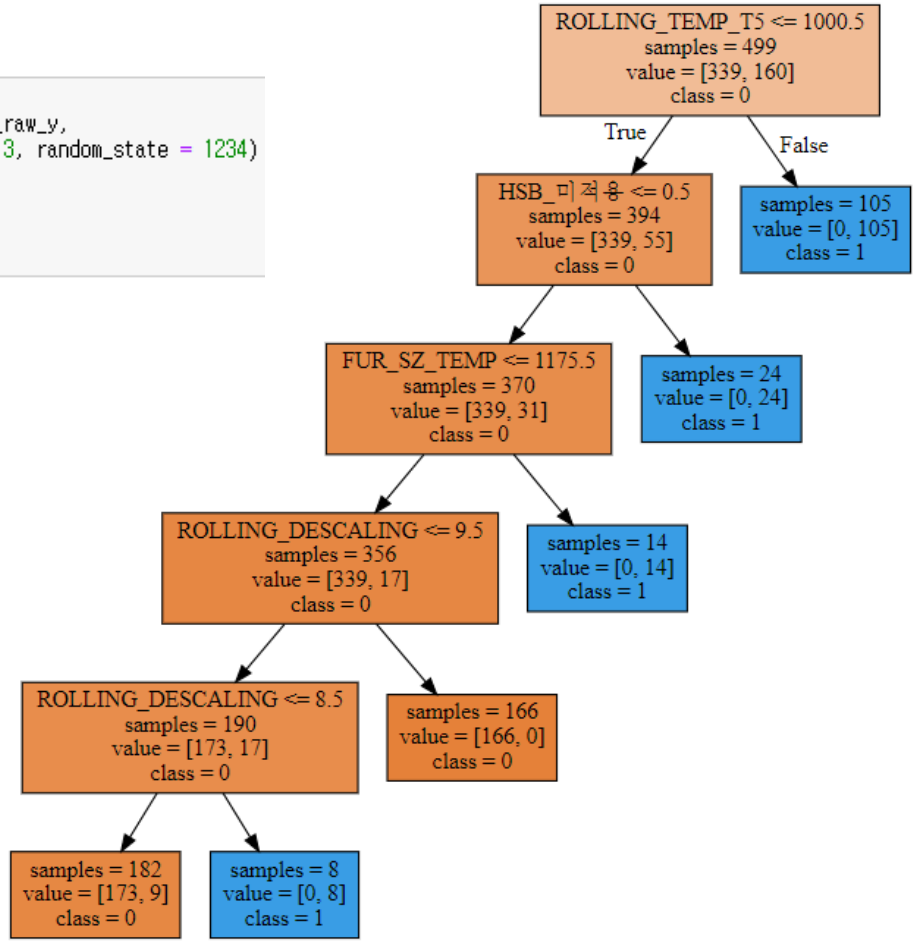
2) 모델 생성

- Grid search 방법을 이용한 결과 max_depth = 5, min_samples_leaf = 6, min_samples_split = 10일 때 가장 좋은 성능을 보임
- Train/test accuracy와 Confusion matrix

```
Train Accuracy: 0.981964

Test Accuracy: 0.986047

Confusion matrix:
[[144   0]
 [  3  68]]
```

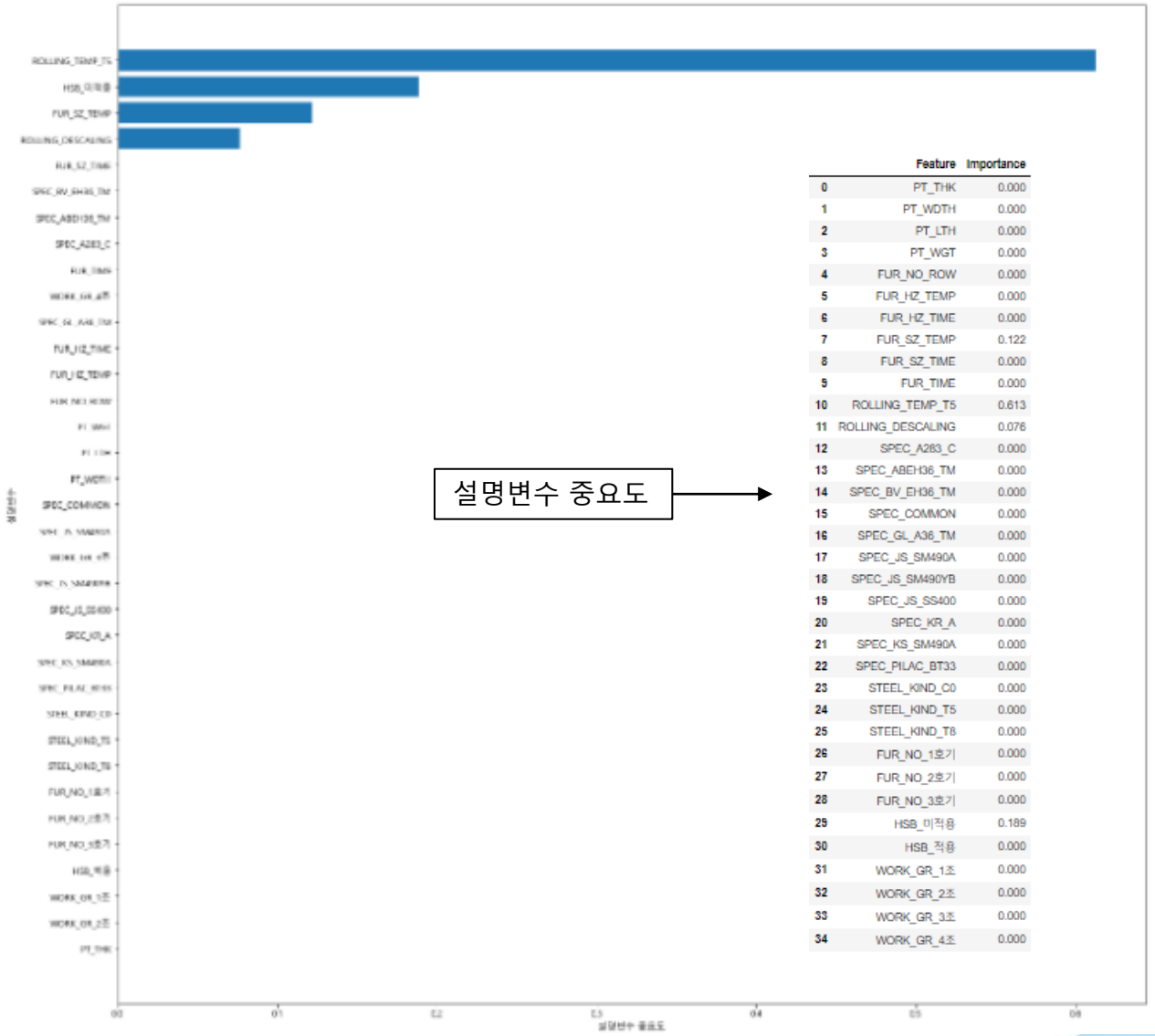


With POSCO

2. 의사결정나무

3) 설명변수 중요도 확인

- ROLLING_TEMP_T5, HSB_미적용, FUR_SZ_TEMP, ROLLING DECSCALING 순으로 영향력이 강함



	Feature	Importance
0	PT_THK	0.000
1	PT_WDTH	0.000
2	PT_LTH	0.000
3	PT_WGT	0.000
4	FUR_NO_ROW	0.000
5	FUR_HZ_TEMP	0.000
6	FUR_HZ_TIME	0.000
7	FUR_SZ_TEMP	0.122
8	FUR_SZ_TIME	0.000
9	FUR_TIME	0.000
10	ROLLING_TEMP_T5	0.613
11	ROLLING_DESCALING	0.076
12	SPEC_A283_C	0.000
13	SPEC_ABEH36_TM	0.000
14	SPEC_BV_EH36_TM	0.000
15	SPEC_COMMON	0.000
16	SPEC_GL_A36_TM	0.000
17	SPEC_JS_SM490A	0.000
18	SPEC_JS_SM490YB	0.000
19	SPEC_JS_SS400	0.000
20	SPEC_KR_A	0.000
21	SPEC_KS_SM490A	0.000
22	SPEC_PILAC_BT33	0.000
23	STEEL_KIND_C0	0.000
24	STEEL_KIND_T5	0.000
25	STEEL_KIND_T8	0.000
26	FUR_NO_1호기	0.000
27	FUR_NO_2호기	0.000
28	FUR_NO_3호기	0.000
29	HSB_미적용	0.189
30	HSB_적용	0.000
31	WORK_GR_1조	0.000
32	WORK_GR_2조	0.000
33	WORK_GR_3조	0.000
34	WORK_GR_4조	0.000

3. 랜덤 포레스트

1) Train/test set 설정

- 설명변수와 목표변수를 나눔
- Train/test set을 7:3 비율로 생성

```
# 데이터들을 train/test data로 분리
df_train_x, df_test_x, df_train_y, df_test_y = train_test_split(df_raw_x, df_raw_y,
                                                    test_size = 0.3, random_state = 1234)
print('train data X size : {}'.format(df_train_x.shape))
print('train data Y size : {}'.format(df_train_y.shape))
print('test data X size : {}'.format(df_test_x.shape))
print('test data Y size : {}'.format(df_test_y.shape))
```

```
train data X size : (499, 35)
train data Y size : (499, 1)
test data X size : (215, 35)
test data Y size : (215, 1)
```

2) 모델 생성

- Grid search 방법을 이용한 결과 n_estimators = 100, max_depth = 7, min_samples_leaf = 2일 때 가장 좋은 성능을 보임
- Train/test accuracy와 Confusion matrix

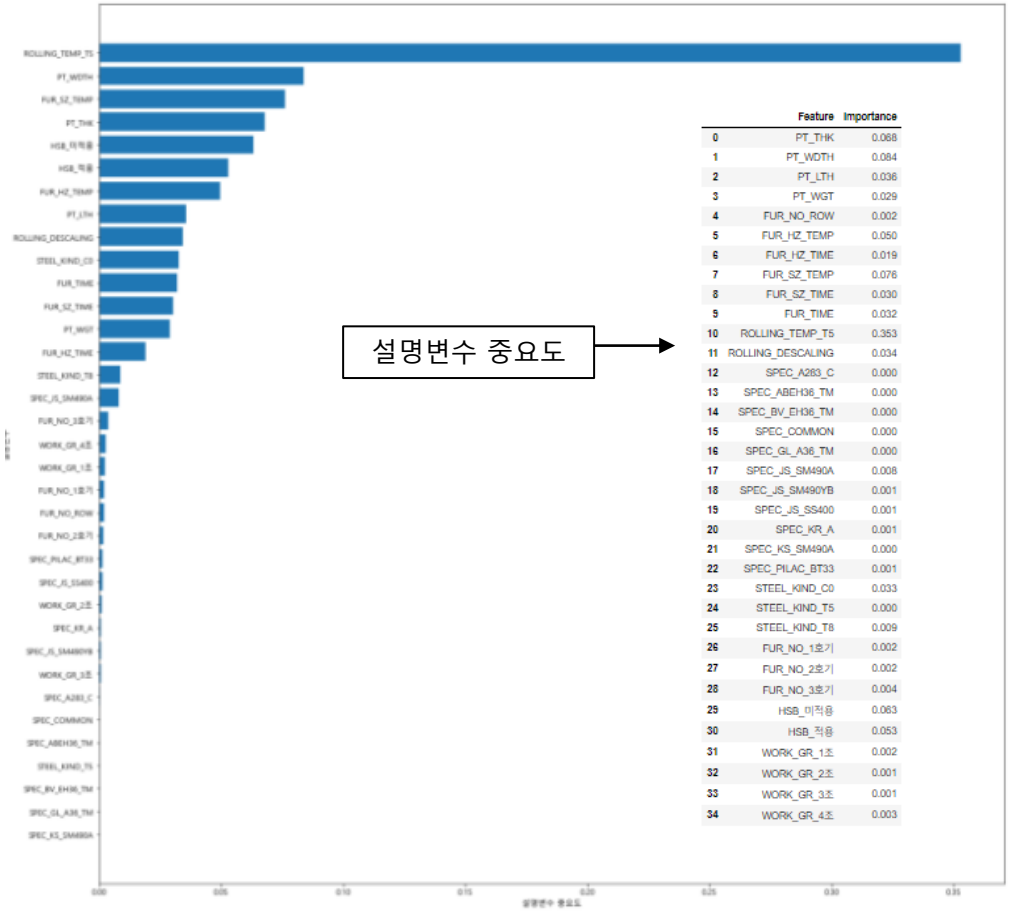
```
Train Accuracy: 0.969940

Test Accuracy: 0.944186

Confusion matrix:
[[144   0]
 [ 12  59]]
```

3) 설명변수 중요도 확인

- ROLLING_TEMP_T5, PT_WIDTH, FUR_SZ_TEMP, PT_THK, HSB(미적용) 순으로 영향력이 강함



	Feature	Importance
0	PT_THK	0.068
1	PT_WDTH	0.084
2	PT_LTH	0.036
3	PT_WGT	0.029
4	FUR_NO_ROW	0.002
5	FUR_HZ_TEMP	0.050
6	FUR_HZ_TIME	0.019
7	FUR_SZ_TEMP	0.076
8	FUR_SZ_TIME	0.030
9	FUR_TIME	0.032
10	ROLLING_TEMP_T5	0.353
11	ROLLING_DESCALING	0.034
12	SPEC_A283_C	0.000
13	SPEC_ABEH36_TM	0.000
14	SPEC_BV_EH36_TM	0.000
15	SPEC_COMMON	0.000
16	SPEC_GL_A36_TM	0.000
17	SPEC_JS_SM490A	0.008
18	SPEC_JS_SM490YB	0.001
19	SPEC_JS_SS400	0.001
20	SPEC_KR_A	0.001
21	SPEC_KS_SM490A	0.000
22	SPEC_PILAC_BT33	0.001
23	STEEL_KIND_C0	0.033
24	STEEL_KIND_T5	0.000
25	STEEL_KIND_T8	0.009
26	FUR_NO_1호기	0.002
27	FUR_NO_2호기	0.002
28	FUR_NO_3호기	0.004
29	HSB_미적용	0.063
30	HSB_적용	0.053
31	WORK_GR_1조	0.002
32	WORK_GR_2조	0.001
33	WORK_GR_3조	0.001
34	WORK_GR_4조	0.003

4. 그래디언트 부스팅

1) Train/test set 설정

- 설명변수와 목표변수를 나눔
- Train/test set을 7:3 비율로 생성

```
# 데이터들을 train/test data로 분리
df_train_x, df_test_x, df_train_y, df_test_y = train_test_split(df_raw_x, df_raw_y,
                                                test_size = 0.3, random_state = 1234)
print('train data X size : {}'.format(df_train_x.shape))
print('train data Y size : {}'.format(df_train_y.shape))
print('test data X size : {}'.format(df_test_x.shape))
print('test data Y size : {}'.format(df_test_y.shape))
```

```
train data X size : (499, 35)
train data Y size : (499, 1)
test data X size : (215, 35)
test data Y size : (215, 1)
```

2) 모델 생성

- Grid search 방법을 이용한 결과 n_estimators = 100, learning_rate = 0.8, max_depth = 4, min_samples_leaf = 10일 때 가장 좋은 성능을 보임
- Train/test accuracy와 Confusion matrix

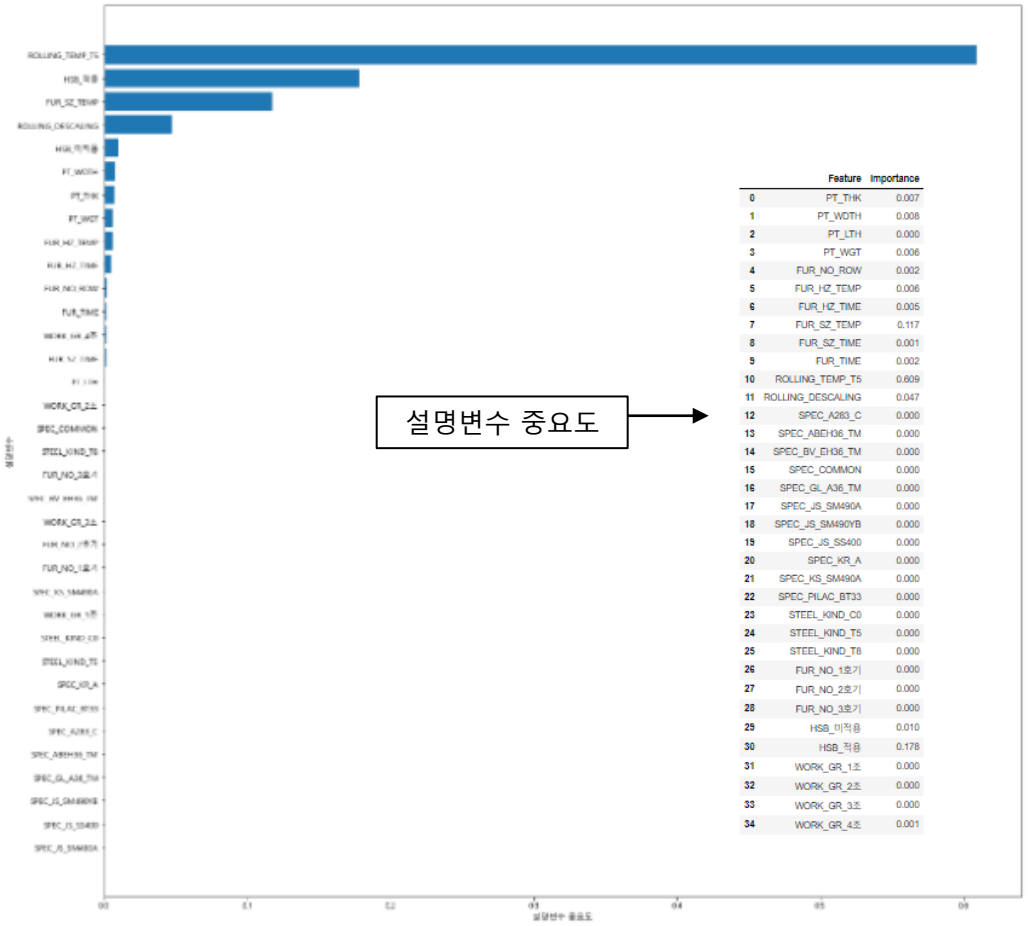
```
Train Accuracy: 1.000000

Test Accuracy: 0.995349

Confusion matrix:
[[143   1]
 [  0  71]]
```

3) 설명변수 중요도 확인

- ROLLING_TEMP_T5, HSB(적용), FUR_SZ_TEMP, ROLLING_DESCALING, HSB(미적용) 순으로 영향력이 강함

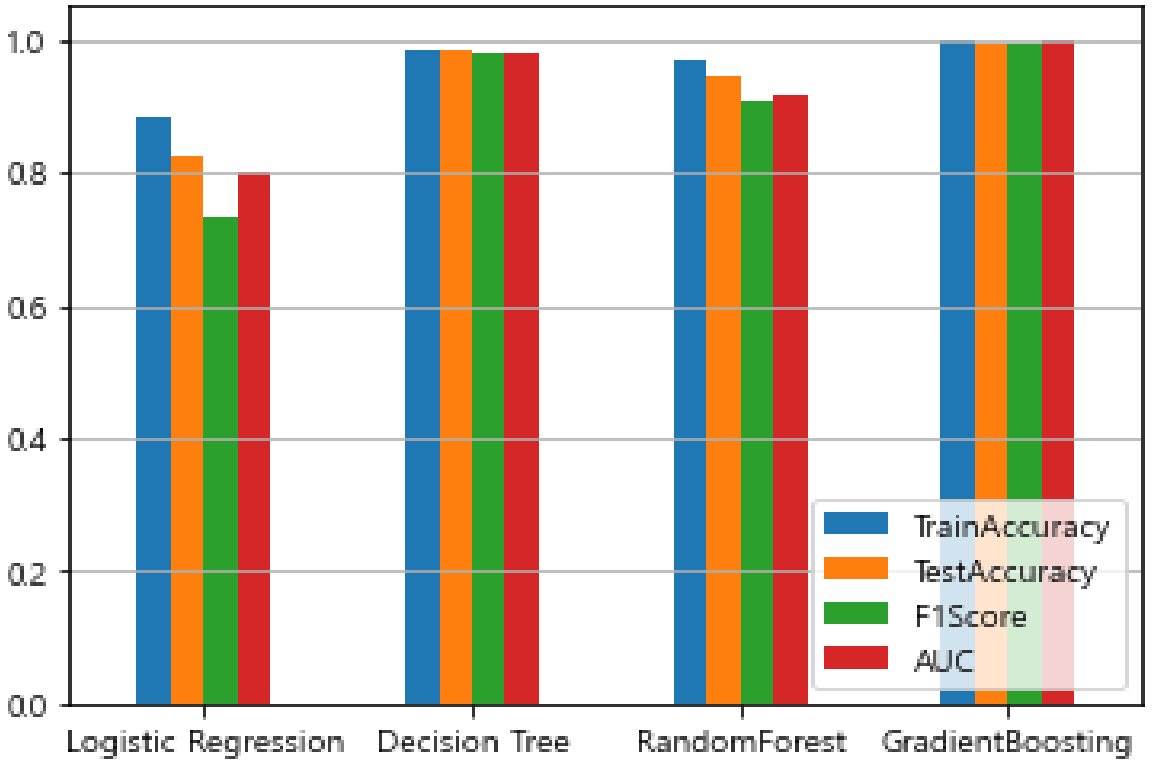


	Feature	Importance
0	PT_THK	0.007
1	PT_WDTH	0.008
2	PT_LTH	0.000
3	PT_WGT	0.006
4	FUR_NO_ROW	0.002
5	FUR_HZ_TEMP	0.006
6	FUR_HZ_TIME	0.005
7	FUR_SZ_TEMP	0.117
8	FUR_SZ_TIME	0.001
9	FUR_TIME	0.002
10	ROLLING_TEMP_T5	0.609
11	ROLLING_DESCALING	0.047
12	SPEC_A283_C	0.000
13	SPEC_ABEH36_TM	0.000
14	SPEC_BV_EH36_TM	0.000
15	SPEC_COMMON	0.000
16	SPEC_GL_A36_TM	0.000
17	SPEC_JS_SM490A	0.000
18	SPEC_JS_SM490YB	0.000
19	SPEC_JS_SS400	0.000
20	SPEC_KR_A	0.000
21	SPEC_KS_SM490A	0.000
22	SPEC_PILAC_BT33	0.000
23	STEEL_KIND_C0	0.000
24	STEEL_KIND_T5	0.000
25	STEEL_KIND_T8	0.000
26	FUR_NO_1호기	0.000
27	FUR_NO_2호기	0.000
28	FUR_NO_3호기	0.000
29	HSB_미적용	0.010
30	HSB_적용	0.178
31	WORK_GR_1조	0.000
32	WORK_GR_2조	0.000
33	WORK_GR_3조	0.000
34	WORK_GR_4조	0.001

5. 생성 모델들의 성능 비교

- 모든 모델들의 train accuracy가 0.8 이상
- 모든 모델들의 test accuracy가 0.8 이상
- 모든 모델들의 F1 score가 0.7 이상
- 모든 모델들의 AUC가 0.8 이상
- 로지스틱 회귀의 경우 모든 평가지표가 최하
- 그래디언트 부스팅의 경우 모든 평가지표가 1등
- 모든 모델들의 평가지표가 나쁘지 않으므로 모델을 통해 분석한 설명변수 중요도를 신뢰할 수 있음

	TrainAccuracy	TestAccuracy	F1Score	AUC
Logistic Regression	0.882	0.823	0.732	0.800
Decision Tree	0.982	0.986	0.978	0.979
RandomForest	0.970	0.944	0.908	0.915
GradientBoosting	1.000	0.995	0.993	0.997



with POSCO

결론 및 대안제시

- **중요인자 선정**

1) 선정 방식

➢ 탐색적 분석 -> 상관성이 있음(1) , 상관성이 없음(0)

➢ 모델링 기법 -> 생성 모델의 설명변수 중요도 확인, 가장 영향을 미치는 5개 변수들을 높은 순부터 5, 4, 3, 2, 1 로 점수 부여

2) 선정 결과

변수 설명	변수 역할	변수 형태	분석 제외 사유	탐색적 기법	모델링 기법				총점	선정
				그래프	로지스틱회귀	의사결정트리	랜덤 포레스트	그래디언트 부스팅		
Plate No	ID	범주형	목표변수와 연관X							
작업시각	제외	연속형	목표변수와 연관X							
Scale불량	목표변수	범주형								
제품 규격	설명변수	범주형		1	0	0	0	0	1	X
강종	설명변수	범주형		1	0	0	0	0	1	X
Plate 뚜께	설명변수	연속형		1	0	0	2	0	3	X
Plate 폭	설명변수	연속형		1	2	0	4	1	8	O
Plate 길이	설명변수	연속형		1	0	0	0	0	1	X
Plate 중량	설명변수	연속형		1	0	0	0	0	1	X
가열로 호기	설명변수	범주형		1	0	0	0	0	1	X
가열로	설명변수	연속형		0	0	0	0	0	0	X
가열로 가열대 온도	설명변수	연속형		1	0	0	0	0	1	X
가열로 가열대 시간	설명변수	연속형		0	5	0	0	0	5	O
가열로 균열대 온도	설명변수	연속형		1	0	3	3	3	10	O
가열로 균열대 시간	설명변수	연속형		1	0	1	0	0	2	X
가열로 시간	설명변수	연속형		1	0	0	0	0	1	X
추출온도	제외	연속형	FUR_SZ_TEMP과 완벽히 겹침							
압연온도	설명변수	연속형		1	4	5	5	5	20	O
HSB적용(1-적용,0-미적용)	설명변수	범주형		1	1	4	1	4	11	O
압연 중 Descaling 횟수	설명변수	연속형		1	3	2	0	2	8	O
작업조	설명변수	범주형		1	0	0	0	0	1	X

➢ 압연온도, HSB, 가열로 균열대 온도, Plate 폭, 압연 중 Descaling 횟수, 가열로 가열대 시간 6개를 중요변수로 선정

with POSCO

결론 및 대안제시

• 결론

- 분석 결과 압연온도, HSB, 가열로 균열대 온도, Plate 폭, 압연 중 Descaling 횟수, 가열로 가열대 시간이 스케일(Scale) 불량에 가장 큰 영향을 미침을 확인
- 중요변수들이 목표변수(SCALE)에 미치는 영향력
  -> 압연온도 > HSB > 가열로 균열대 온도 > Plate 폭 > 압연 중 Descaling 횟수 > 가열로 가열대 시간
- 탐색적 분석을 통한 중요변수들과 설명변수의 상관관계

스케일(Scale)	압연온도	HSB	가열로 균열대 온도	Plate 폭	압연 중 Descaling 횟수	가열로 가열대 시간
없음 ↕ 발생	저 ↕ 고	적용 ↕ 미적용	저 ↕ 고	두꺼움 ↕ 얇음	나머지 ↕ 8	많음 ↕ 적음

• 대안 제시

- 분석 결과 압연온도가 목표변수에 가장 영향을 미치므로 압연온도를 낮춰 스케일 불량을 줄인다.
- HSB를 미적용했을 경우 모두 Scale 불량이 나왔으므로 압연 공정에 HSB 적용을 필수로 한다.
- Plate 폭을 얇게 할 때는 스케일 불량 발생률이 높으므로 주의한다.
- 압연 중 Descaling 횟수는 가능하면 적은 횟수로 한다.
- 가열로 가열대 시간은 충분히 오래 한다.

느낀 점 및 소감

- 압연 공정과 스케일(Scale) 발생 원리에 대해 알게 됨
- 스케일 발생을 일으키는 중요인자들을 알 수 있었음
- 로지스틱 회귀, 의사결정나무, 랜덤 포레스트, 그래디언트 부스팅 기법을 통해 목표변수를 예측하고 실제값과 비교해볼수 있었음
- 중요인자들을 통해 대응방안을 수립 및 제안할 수 있었음