

압연공정 불량 발생 원인 분석

성수호

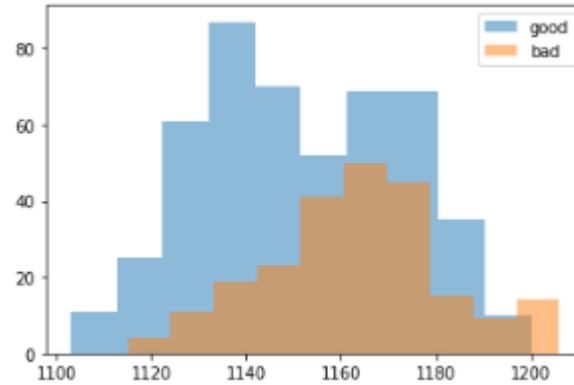
탐색적 분석 (그래프 분석)

압연공정 불량 발생 원인 분석

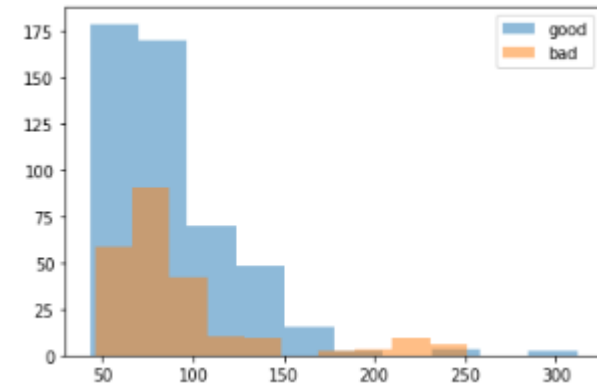
| STEEL_KIND | C0 | C1 | C3 | T0 | T1 | T3 | T5 | T7 | T8 |
|------------|-----|----|----|----|----|----|----|----|----|
| SCALE | | | | | | | | | |
| 불량 | 212 | 1 | 1 | 2 | 2 | 0 | 2 | 6 | 5 |
| 양품 | 291 | 0 | 6 | 14 | 16 | 2 | 41 | 29 | 90 |

| STEEL_KIND | C0 | C1 | C3 | T0 | T1 | T3 | T5 | T7 | T8 |
|------------|-------|-----|-------|-------|-------|-----|-------|-------|-------|
| SCALE | | | | | | | | | |
| 불량 | 0.421 | 1.0 | 0.143 | 0.125 | 0.111 | 0.0 | 0.047 | 0.171 | 0.053 |
| 양품 | 0.579 | 0.0 | 0.857 | 0.875 | 0.889 | 1.0 | 0.953 | 0.829 | 0.947 |

STEEL_KIND



FUR_HZ_TEMP



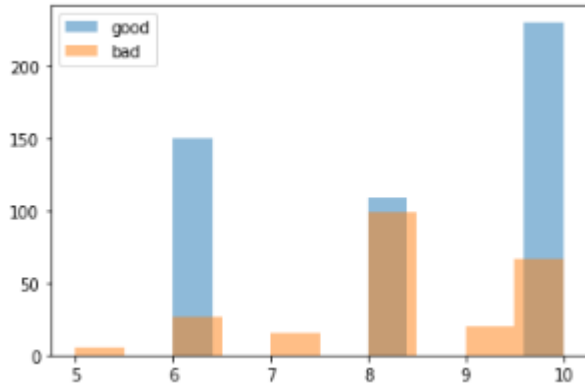
FUR_HZ_TIME

Check POINT

- ✓ STEEL_KIND : 특정 범주에서 불균형한 불량률을 보이면서 종류에 따라 불량에 영향을 미칠 수 있을 것이다.
- ✓ FUR_HZ_TEMP : 가열대의 온도가 높아질수록 불량품 발생 빈도수가 급격하게 높아진다. 온도에 따른 가열대에서 문제가 발생할 수 있으므로 분석이 필요하다.
- ✓ FUR_HZ_TIME : 작업 초반의 경우 불량률이 평균에 유사하지만 특정 시간대에서 불규칙한 비율을 보인다.

탐색적 분석 (그래프 분석)

압연공정 불량 발생 원인 분석

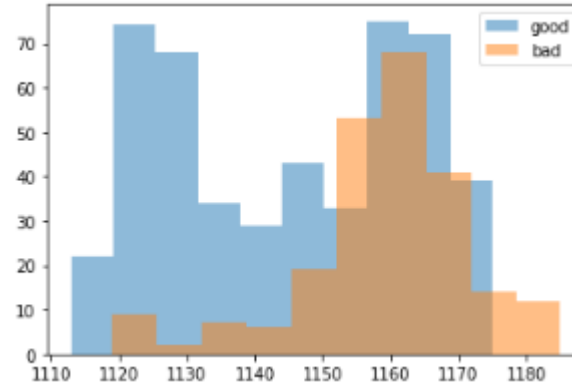


ROLLING_DESCALING

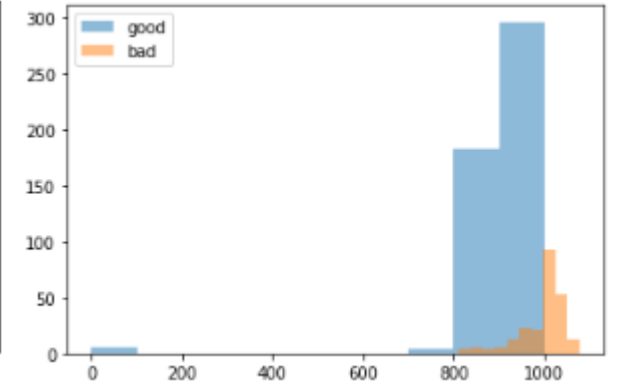
| | | |
|-------|-----|-----|
| HSB | 미적용 | 적용 |
| SCALE | | |
| 불량 | 33 | 198 |
| 양품 | 0 | 489 |

| | | |
|-------|-----|-------|
| HSB | 미적용 | 적용 |
| SCALE | | |
| 불량 | 1.0 | 0.288 |
| 양품 | 0.0 | 0.712 |

HSB



FUR_EXTEMP



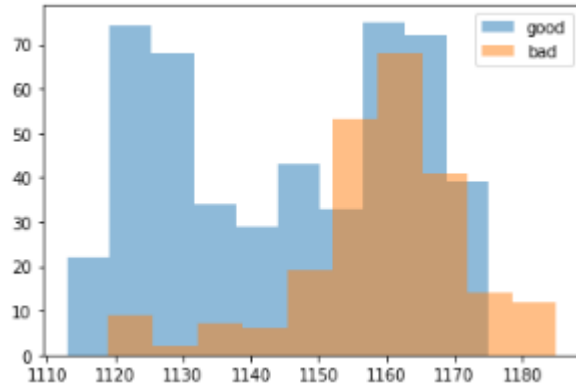
ROLLING_TEMP_T5

Check POINT

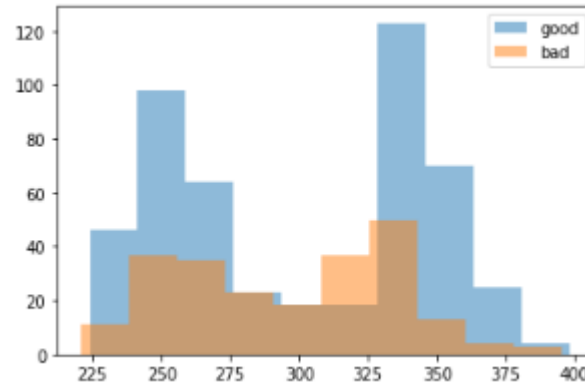
- ✓ ROLLING_DESCALING : 5,7회에서 불량품만 나오며 전체적으로 불량률이 평균보다 높기 때문에 조사 필요.
- ✓ HSB : HSB를 미적용하면 불량만 발생하므로 분석이 필요.
- ✓ FUR_EXTEMP : 추출온도 범위 내에서 불량품 발생률이 극단적으로 차이가 나므로 설명변수로 채택한다.
- ✓ ROLLING_TEMP_T5 : 압연온도는 800도 부근에서 작업이 시작. 0도에서 발생한 데이터는 이상치로 분류.

탐색적 분석 (그래프 분석)

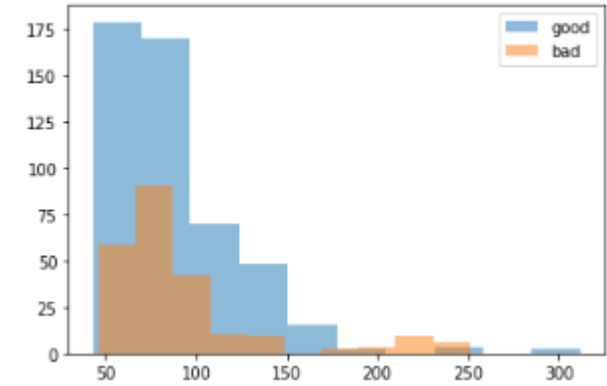
압연공정 불량 발생 원인 분석



FUR_SZ_TEMP



FUR_SZ_TIME



FUR_TIME

Check POINT

- ✓ FUR_SZ_TEMP : 균열대 온도 범위 내에서 불량품 발생률이 극단적으로 차이가 나므로 설명변수로 채택한다.
- ✓ FUR_SZ_TIME : : 중간 시간대에서 양품과 불량품의 발생 빈도가 같으므로 분석이 필요.
- ✓ FUR_TIME : 가열로 시간이 짧을 수록 불량률이 규칙적이거나 150 이상이 되면 불량률이 50%를 초과한다.

탐색적 분석 (그래프 분석)

압연공정 불량 발생 원인 분석

| | | | | | | | | | |
|-------|-------------|--------|---------|---------|------|---------|------|------------|---|
| SPEC | A131-DH36TM | A283-C | A516-60 | A709-36 | AB/A | AB/AH32 | AB/B | AB/EH32-TM | \ |
| SCALE | | | | | | | | | |
| 불량 | 1 | 5 | 1 | 0 | 4 | 1 | 3 | 0 | |
| 양품 | 0 | 1 | 1 | 1 | 3 | 4 | 3 | 2 | |

| | | | | | | | | |
|-------|------------|------------|-----|-----------|-----------|------|-----------|---|
| SPEC | AB/EH36-TM | API-2W-50T | ... | NV-A32-TM | NV-A36-TM | NV-B | NV-D32-TM | \ |
| SCALE | | | | | | | | |
| 불량 | 1 | 0 | ... | 1 | 0 | 2 | 0 | |
| 양품 | 16 | 2 | ... | 2 | 2 | 1 | 4 | |

| | | | | | | | |
|-------|-----------|-----------|-----------|------------|---------|---------|--|
| SPEC | NV-D36-TM | NV-E32-TM | NV-E36-TM | PILAC-BT33 | SA283-C | V42JBN3 | |
| SCALE | | | | | | | |
| 불량 | 1 | 0 | 0 | 2 | 10 | 1 | |
| 양품 | 4 | 2 | 5 | 38 | 11 | 3 | |

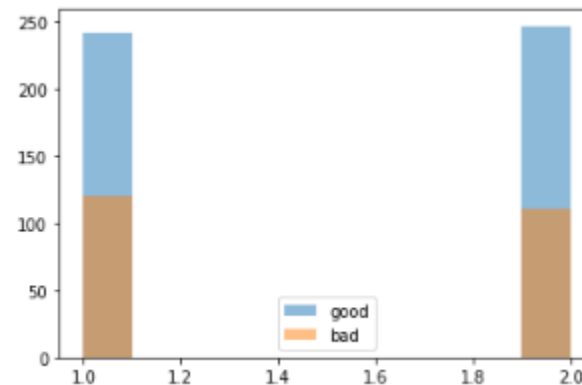
[2 rows x 66 columns]

SPEC

| | | | |
|--------|-----|-----|-----|
| FUR_NO | 1호기 | 2호기 | 3호기 |
| SCALE | | | |
| 불량 | 73 | 70 | 88 |
| 양품 | 167 | 167 | 155 |

| | | | |
|--------|-------|-------|-------|
| FUR_NO | 1호기 | 2호기 | 3호기 |
| SCALE | | | |
| 불량 | 0.304 | 0.295 | 0.362 |
| 양품 | 0.696 | 0.705 | 0.638 |

FUR_NO



FUR_NO_ROW

| | | | | |
|---------|-----|-----|-----|-----|
| WORK_GR | 1조 | 2조 | 3조 | 4조 |
| SCALE | | | | |
| 불량 | 67 | 45 | 54 | 65 |
| 양품 | 122 | 120 | 118 | 129 |

| | | | | |
|---------|-------|-------|-------|-------|
| WORK_GR | 1조 | 2조 | 3조 | 4조 |
| SCALE | | | | |
| 불량 | 0.354 | 0.273 | 0.314 | 0.335 |
| 양품 | 0.646 | 0.727 | 0.686 | 0.665 |

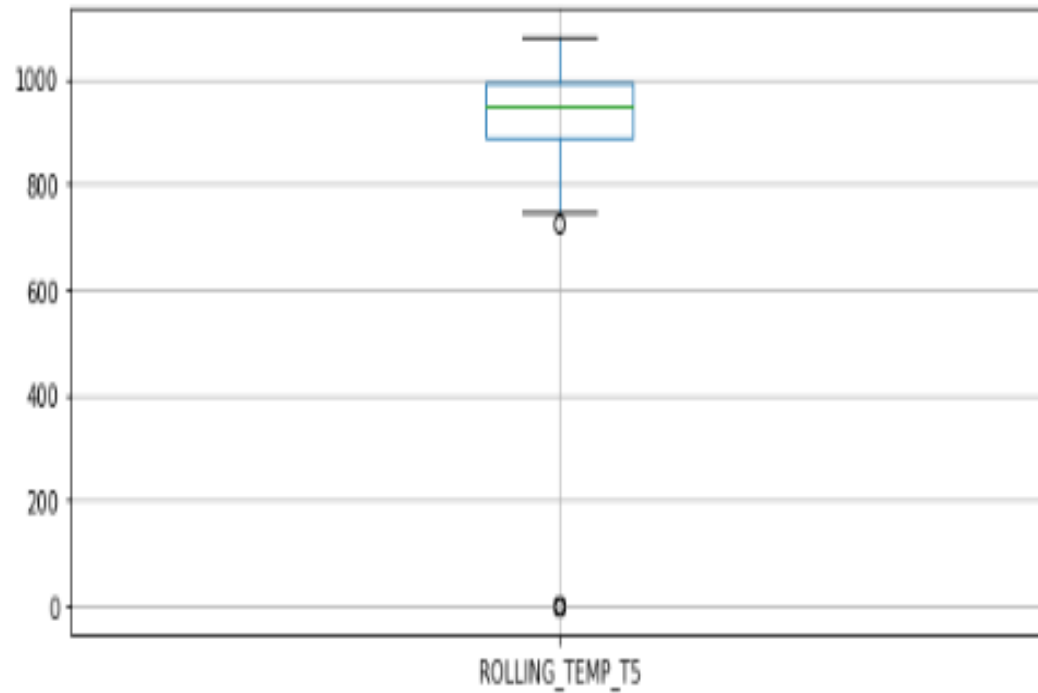
WORK_GR

Check POINT

- ✓ SPEC : 범주가 너무 많아서 위 자료로 유의미성을 파악하기 어렵다. 새로운 범주로 만들어 분석할 필요가 있다.
- ✓ FUR_NO : 가열로의 전 호기에 대해 불량률이 유사하므로 변수에서 제외한다.
- ✓ FUR_NO_ROW : 가열로의 작업순번과 상관없이 불량률이 유사하므로 변수에서 제외한다.
- ✓ WORK_GR : 불량률이 전체적으로 0.3에 근사하기 때문에 작업조 데이터는 크게 유효하지 않을 것이다.
- ✓ 제품의 사이즈와 관련된 변수들은 주문자 생산방식을 고려하여 사이즈 변경이 불가하다 판단했고, 그래서 변수에서 제외 하였다.

탐색적 분석 (이상치 제거)

압연공정 불량 발생 원인 분석



```
1 q1 = np.percentile(df_raw['ROLLING_TEMP_T5'],25)
2 q3 = np.percentile(df_raw['ROLLING_TEMP_T5'],75)
3 minimum=q1-1.5*(q3-q1)
4 df_raw=df_raw[df_raw['ROLLING_TEMP_T5']>=float(minimum)]
5 df_raw['ROLLING_TEMP_T5'].describe()
6
```

```
count    713.000000
mean     942.075736
std       65.344140
min       745.000000
25%      891.000000
50%      952.000000
75%      995.000000
max      1078.000000
Name: ROLLING_TEMP_T5, dtype: float64
```

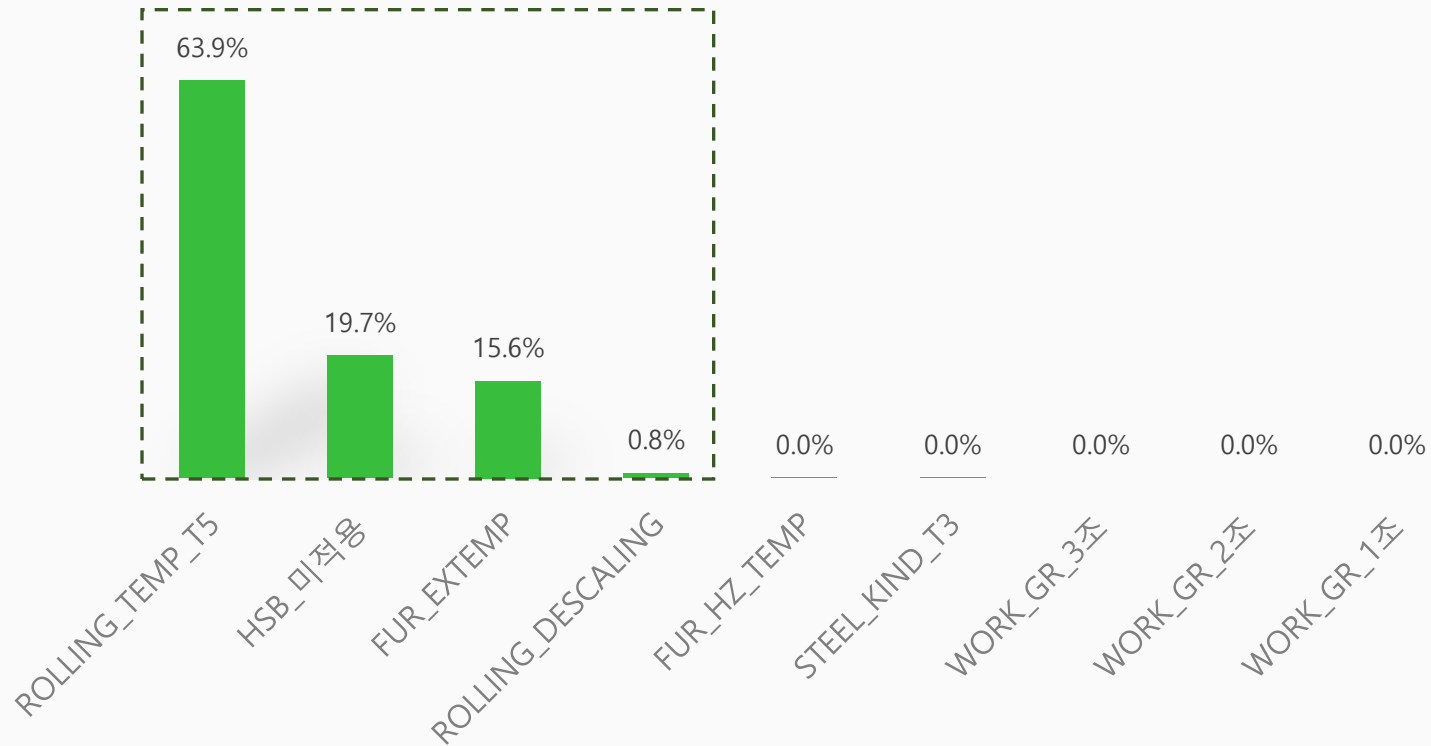
Check POINT

- ✓ 압연온도가 보통 800도 쯤에서 작업이 시작되는데 0도에서는 가열이 되지 않는 것이 당연하므로 이상치로 판단하고 제거한다.
- ✓ 사분위수를 계산하여 최소값 보다 작은 이상치를 제거한다.

탐색적 분석 (변수 중요도 파악)

압연공정 불량 발생 원인 분석

변수 중요도



| | Feature | Importance |
|----|-------------------|------------|
| 6 | ROLLING_TEMP_T5 | 0.639 |
| 17 | HSB_미적용 | 0.197 |
| 5 | FUR_EXTEMP | 0.156 |
| 7 | ROLLING_DESCALING | 0.008 |
| 0 | FUR_HZ_TEMP | 0.000 |
| 13 | STEEL_KIND_T3 | 0.000 |
| 21 | WORK_GR_3조 | 0.000 |
| 20 | WORK_GR_2조 | 0.000 |
| 19 | WORK_GR_1조 | 0.000 |
| 18 | HSB_적용 | 0.000 |
| 16 | STEEL_KIND_T8 | 0.000 |
| 15 | STEEL_KIND_T7 | 0.000 |
| 14 | STEEL_KIND_T5 | 0.000 |
| 11 | STEEL_KIND_T0 | 0.000 |
| 12 | STEEL_KIND_T1 | 0.000 |
| 1 | FUR_HZ_TIME | 0.000 |
| 10 | STEEL_KIND_C3 | 0.000 |
| 9 | STEEL_KIND_C1 | 0.000 |
| 8 | STEEL_KIND_C0 | 0.000 |
| 4 | FUR_TIME | 0.000 |
| 3 | FUR_SZ_TIME | 0.000 |
| 2 | FUR_SZ_TEMP | 0.000 |
| 22 | WORK_GR_4조 | 0.000 |

Check POINT

✓ Decision Tree에서 변수 중요도를 계산하여 위와 같은 변수 순서로 설명력이 높다는 것을 알았다.

탐색적 분석 (로지스틱 회귀분석)

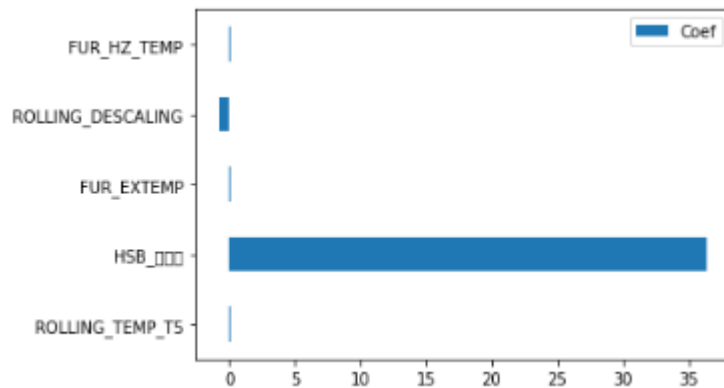
압연공정 불량 발생 원인 분석

Warning: Maximum number of iterations has been exceeded.
Current function value: 0.293118
Iterations: 35

Logit Regression Results

| | | | |
|------------------|------------------|-------------------|-----------|
| Dep. Variable: | SCALE | No. Observations: | 499 |
| Model: | Logit | Df Residuals: | 493 |
| Method: | MLE | Df Model: | 5 |
| Date: | Tue, 14 Jul 2020 | Pseudo R-squ.: | 0.5371 |
| Time: | 22:54:13 | Log-Likelihood: | -146.27 |
| converged: | False | LL-Null: | -315.98 |
| Covariance Type: | nonrobust | LLR p-value: | 3.308e-71 |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|-------------------|----------|----------|----------|-------|-----------|----------|
| Intercept | -98.4971 | 16.601 | -5.933 | 0.000 | -131.034 | -65.960 |
| ROLLING_TEMP_T5 | 0.0371 | 0.005 | 8.144 | 0.000 | 0.028 | 0.046 |
| HSB 미적용 | 36.3964 | 2.37e+06 | 1.54e-05 | 1.000 | -4.64e+06 | 4.64e+06 |
| FUR_EXTTEMP | 0.0291 | 0.023 | 1.274 | 0.203 | -0.016 | 0.074 |
| ROLLING_DESCALING | -0.8108 | 0.151 | -5.377 | 0.000 | -1.106 | -0.515 |
| FUR_HZ_TEMP | 0.0303 | 0.014 | 2.209 | 0.027 | 0.003 | 0.057 |



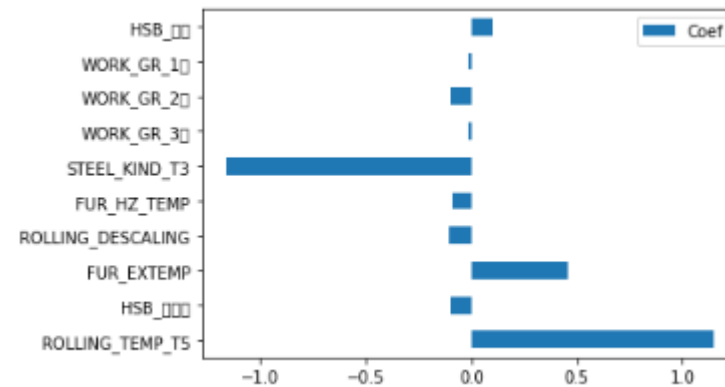
scaling 전

Warning: Maximum number of iterations has been exceeded.
Current function value: 0.495405
Iterations: 35

Logit Regression Results

| | | | |
|------------------|------------------|-------------------|-----------|
| Dep. Variable: | SCALE | No. Observations: | 491 |
| Model: | Logit | Df Residuals: | 481 |
| Method: | MLE | Df Model: | 9 |
| Date: | Tue, 14 Jul 2020 | Pseudo R-squ.: | 0.2133 |
| Time: | 23:00:03 | Log-Likelihood: | -243.24 |
| converged: | False | LL-Null: | -309.19 |
| Covariance Type: | nonrobust | LLR p-value: | 4.824e-24 |

| | coef | std err | z | P> z | [0.025 | 0.975] |
|-------------------|---------|---------|--------|-------|--------|--------|
| Intercept | -1.1052 | 0.406 | -2.721 | 0.007 | -1.901 | -0.309 |
| ROLLING_TEMP_T5 | 1.1495 | 0.103 | 6.268 | 0.000 | 0.790 | 1.509 |
| HSB 미적용 | -0.0984 | nan | nan | nan | nan | nan |
| FUR_EXTTEMP | 0.4529 | 0.284 | 1.592 | 0.111 | -0.105 | 1.010 |
| ROLLING_DESCALING | -0.1094 | 0.161 | -0.681 | 0.496 | -0.424 | 0.206 |
| FUR_HZ_TEMP | -0.0921 | 0.201 | -0.458 | 0.647 | -0.486 | 0.302 |
| STEEL_KIND_T3 | -1.1643 | 3.593 | -0.324 | 0.746 | -8.206 | 5.878 |
| WORK_GR_3조 | -0.0112 | 0.132 | -0.085 | 0.932 | -0.269 | 0.247 |
| WORK_GR_2조 | -0.0952 | 0.132 | -0.721 | 0.471 | -0.354 | 0.163 |
| WORK_GR_1조 | -0.0147 | 0.136 | -0.109 | 0.914 | -0.280 | 0.251 |
| HSB 적용 | 0.0984 | nan | nan | nan | nan | nan |



scaling 후

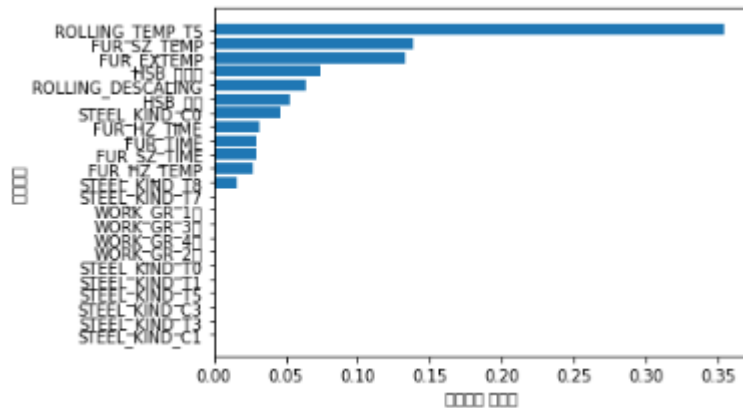
Check POINT

- ✓ 설명 변수가 얼마나 영향이 있는지 서로 비교하기 위해 표준화(scaling)을 했다.
- ✓ 로지스틱 회귀분석에서는 ROLLING_TEMP_T5의 값이 커질수록 불량률의 영향이 커진다.
- ✓ 강종(STEEL KIND)가 T3일 경우 불량률이 낮아질 가능성이 커진다.

모델링 1. 랜덤포레스트

압연공정 불량 발생 원인 분석

| | Feature | Importance |
|----|-------------------|------------|
| 6 | ROLLING_TEMP_T5 | 0.355 |
| 2 | FUR_SZ_TEMP | 0.138 |
| 5 | FUR_EXTTEMP | 0.133 |
| 17 | HSB_미적용 | 0.074 |
| 7 | ROLLING_DESCALING | 0.064 |
| 18 | HSB_적용 | 0.053 |
| 8 | STEEL_KIND_C0 | 0.046 |
| 1 | FUR_HZ_TIME | 0.031 |
| 4 | FUR_TIME | 0.029 |
| 3 | FUR_SZ_TIME | 0.029 |
| 0 | FUR_HZ_TEMP | 0.028 |
| 16 | STEEL_KIND_T8 | 0.016 |
| 15 | STEEL_KIND_T7 | 0.001 |
| 19 | WORK_GR_1조 | 0.001 |
| 21 | WORK_GR_3조 | 0.000 |
| 22 | WORK_GR_4조 | 0.000 |
| 20 | WORK_GR_2조 | 0.000 |
| 11 | STEEL_KIND_T0 | 0.000 |
| 12 | STEEL_KIND_T1 | 0.000 |
| 14 | STEEL_KIND_T5 | 0.000 |
| 13 | STEEL_KIND_T3 | 0.000 |
| 10 | STEEL_KIND_C3 | 0.000 |
| 9 | STEEL_KIND_C1 | 0.000 |



```
Accuracy on training set:0.954
Accuracy on test set:0.935
Confusion Matrix:
[[147  0]
 [ 14 53]]
```

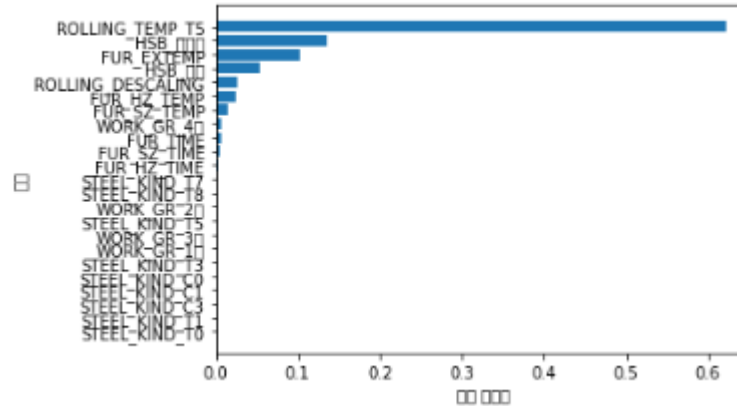
Check POINT

- ✓ 트리 생성 시 압연온도, 균열대 온도 순으로 영향도가 크다고 해석할 수 있다.
- ✓ 변수 중요도는 낮더라도 훈련 데이터에 따라 중요도가 변경될 수 있다.

모델링 2. 그래디언트 부스팅

압연공정 불량 발생 원인 분석

| | Feature | Importance |
|----|-------------------|------------|
| 6 | ROLLING_TEMP_T5 | 0.768 |
| 0 | FUR_HZ_TEMP | 0.052 |
| 2 | FUR_SZ_TEMP | 0.048 |
| 5 | FUR_EXTEMP | 0.042 |
| 7 | ROLLING_DESCALING | 0.035 |
| 4 | FUR_TIME | 0.018 |
| 3 | FUR_SZ_TIME | 0.013 |
| 1 | FUR_HZ_TIME | 0.011 |
| 22 | WORK_GR_4조 | 0.008 |
| 19 | WORK_GR_1조 | 0.005 |
| 16 | STEEL_KIND_T8 | 0.001 |
| 21 | WORK_GR_3조 | 0.000 |
| 9 | STEEL_KIND_C1 | 0.000 |
| 10 | STEEL_KIND_C3 | 0.000 |
| 8 | STEEL_KIND_C0 | 0.000 |
| 12 | STEEL_KIND_T1 | 0.000 |
| 13 | STEEL_KIND_T3 | 0.000 |
| 14 | STEEL_KIND_T5 | 0.000 |
| 15 | STEEL_KIND_T7 | 0.000 |
| 17 | HSB_미적용 | 0.000 |
| 18 | HSB_적용 | 0.000 |
| 20 | WORK_GR_2조 | 0.000 |
| 11 | STEEL_KIND_T0 | 0.000 |



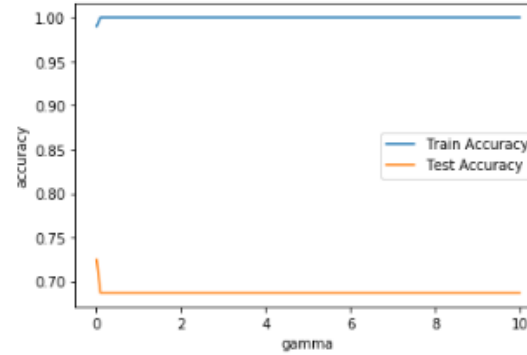
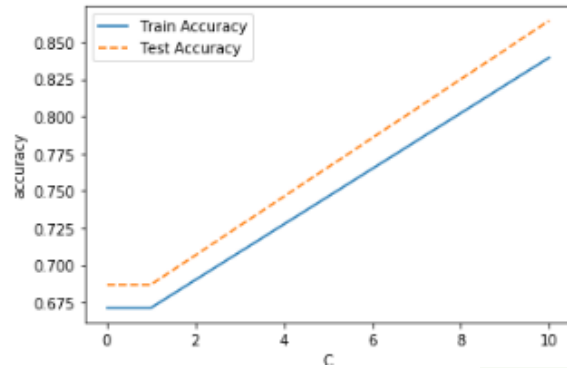
```
Accuracy on training set:0.944
Accuracy on test set:0.935
Confusion matrix:
[[146   1]
 [ 13  54]]
```

Check POINT

- ✓ 트리 생성 시 랜덤포레스트와 유사하게 압연온도, 균열대 온도 순으로 영향도를 보여주고 있다.
- ✓ 변수 중요도는 낮더라도 훈련 데이터에 따라 중요도가 변경될 수 있다.

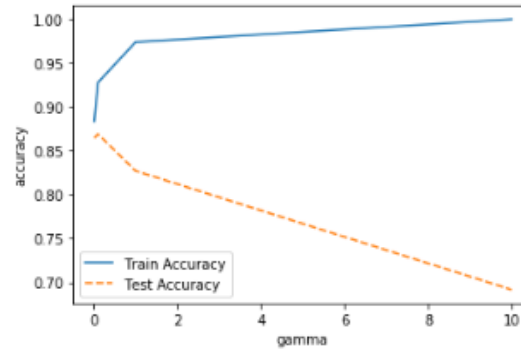
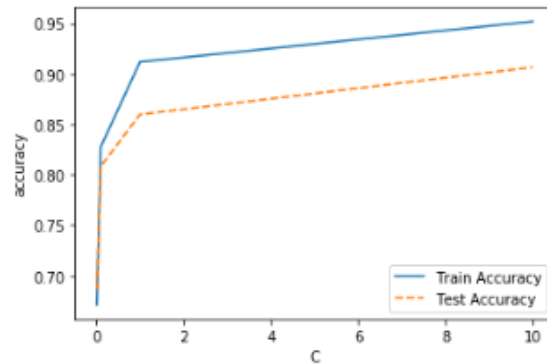
모델링 3. SVM

압연공정 불량 발생 원인 분석



Accuracy on training set:0.671
Accuracy on test set:0.687

SCALING 전, 후
C, gamma 변화에 따른
정확도



정확도 향상

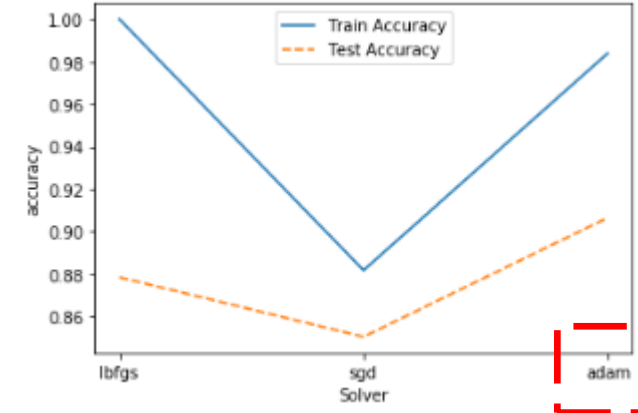
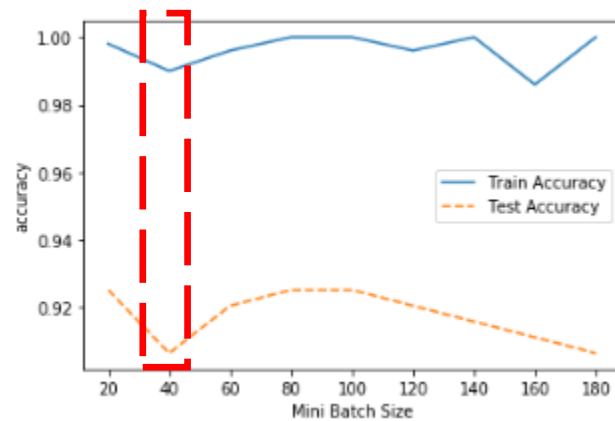
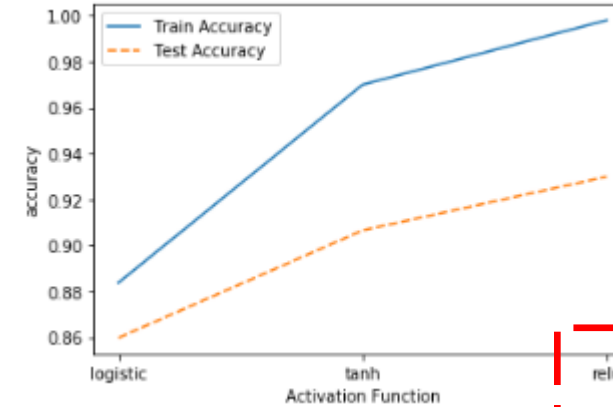
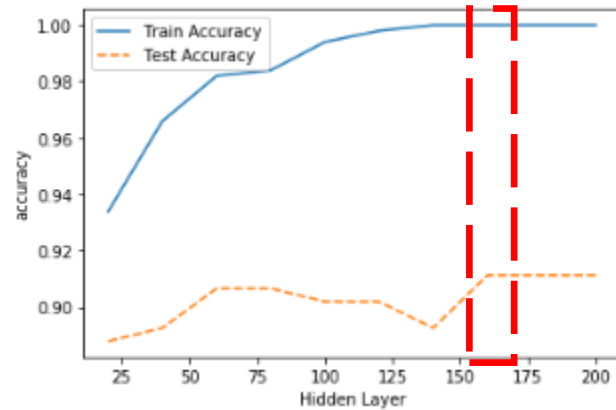
Accuracy on training set:0.912
Accuracy on test set:0.860

Check POINT

✓ 적절한 C와 gamma를 찾기 위해 scaling이 선행 되었고, 그 결과로 높은 정확도의 모델 형성

모델링 4. 인공지능경망

압연공정 불량 발생 원인 분석

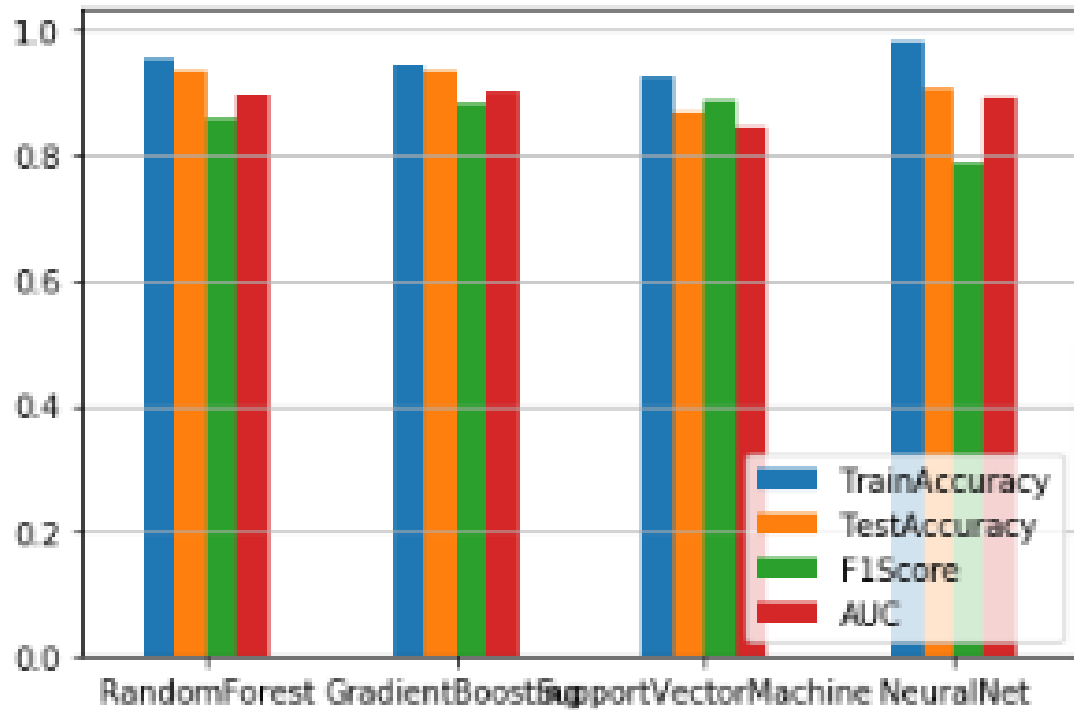


Check POINT

- ✓ 2개의 은닉층에 160개의 노드를 넣어 오차 최소화
- ✓ 활성화 함수로는 정확도가 가장 높은 relu를 선택했다.
- ✓ adam이 test에서 실제값과 예측값의 차이를 가장 최소화 한다.
- ✓ 전체 학습 데이터를 batch 사이즈로 등분하여 test의 오차를 줄인다.

모델 평가

압연공정 불량 발생 원인 분석



| | TrainAccuracy | TestAccuracy | F1Score | AUC |
|----------------------|---------------|--------------|---------|-------|
| RandomForest | 0.954 | 0.935 | 0.857 | 0.896 |
| GradientBoosting | 0.944 | 0.935 | 0.883 | 0.900 |
| SupportVectorMachine | 0.928 | 0.869 | 0.885 | 0.844 |
| NeuralNet | 0.984 | 0.907 | 0.788 | 0.891 |

AUC(AREA UNDER COVER)

: 1에 가까울 수록 완벽한 모델

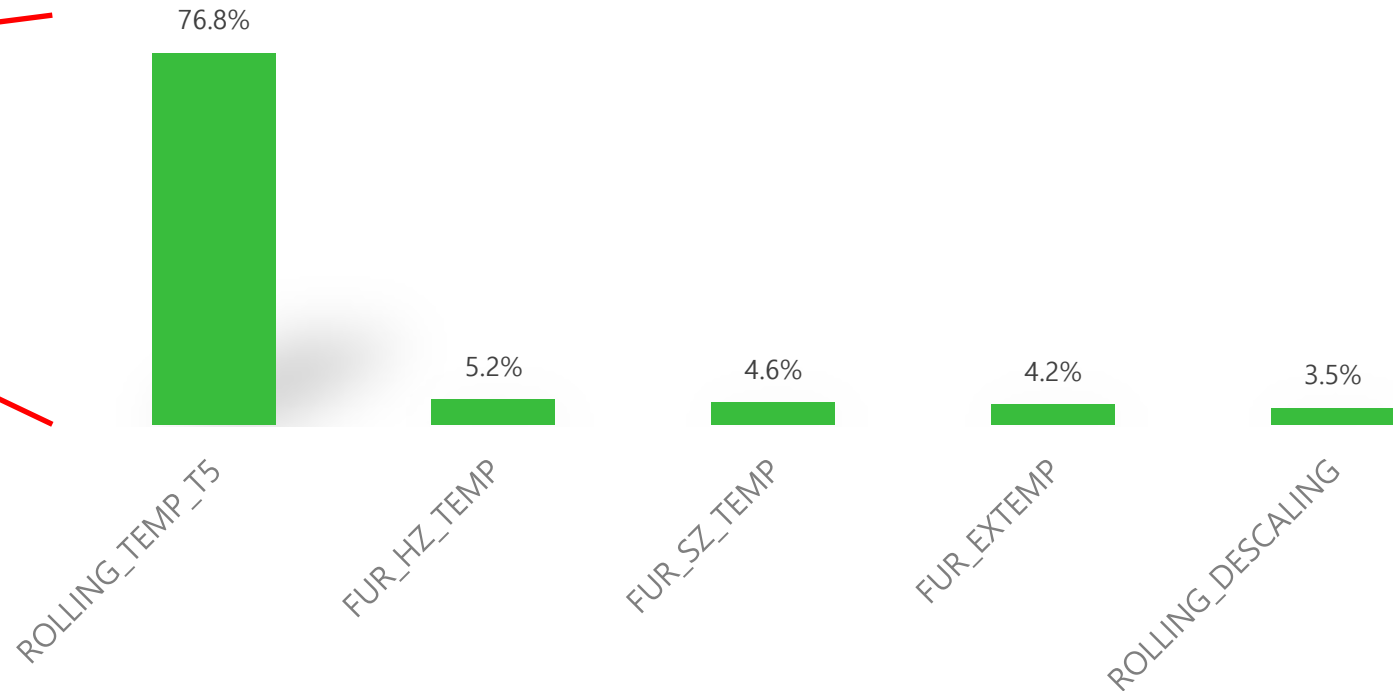
Check POINT

- ✓ train과 test에 대한 정확도 차이를 각 모델별 비교했을 때 인공신경망의 train 정확도가 가장 높지만 test의 정확도가 너무 낮아 과적합이 예상된다.
- ✓ 그래디언트 부스팅은 테스트에 가장 적합한 모델이며 AUC도 가장 높게 나왔다.
- ✓ 인공신경망은 낮은 F1 score를 보여주는데 이는 모델의 정밀도 혹은 재현율의 균형이 제일 안 좋다는 의미.

결과 해석

압연공정 불량 발생 원인 분석

| | Feature | Importance |
|----|-------------------|------------|
| 6 | ROLLING_TEMP_T5 | 0.768 |
| 0 | FUR_HZ_TEMP | 0.052 |
| 2 | FUR_SZ_TEMP | 0.048 |
| 5 | FUR_EXTEMP | 0.042 |
| 7 | ROLLING_DESCALING | 0.035 |
| 4 | FUR_TIME | 0.018 |
| 3 | FUR_SZ_TIME | 0.013 |
| 1 | FUR_HZ_TIME | 0.011 |
| 22 | WORK_GR_4조 | 0.008 |
| 19 | WORK_GR_1조 | 0.005 |
| 16 | STEEL_KIND_T8 | 0.001 |
| 21 | WORK_GR_3조 | 0.000 |
| 9 | STEEL_KIND_C1 | 0.000 |
| 10 | STEEL_KIND_C3 | 0.000 |
| 8 | STEEL_KIND_C0 | 0.000 |
| 12 | STEEL_KIND_T1 | 0.000 |
| 13 | STEEL_KIND_T3 | 0.000 |
| 14 | STEEL_KIND_T5 | 0.000 |
| 15 | STEEL_KIND_T7 | 0.000 |
| 17 | HSB_미적용 | 0.000 |
| 18 | HSB_적용 | 0.000 |
| 20 | WORK_GR_2조 | 0.000 |
| 11 | STEEL_KIND_T0 | 0.000 |



결과

- ✓ 모델 평가에서 가장 높은 정확도를 보여준 그래디언트 부스팅 모델의 변수 중요도의 차트이다.
- ✓ 사전 탐색적 분석에서 이뤄진 변수 중요도와 비교했을 때 60% 정도가 일치한다.
- ✓ 불량 발생의 가장 큰 원인은 압연온도이며 모든 모델에서 가장 큰 원인으로 분석됐다.
- ✓ 5가지 요인 중 1위에서 순서대로 모두 온도와 관련돼 있으며 공정과정 중 온도가 높아 질수록 불량 발생률이 증가했다. 온도에 대한 설정과 고온 발생시 공정에 발생하는 문제에 대한 구체적인 조사가 필요할 것이라 생각한다.