posco

후판공정 불량예측 및 개선방안

청년 AI·Big Data 아카데미 25기

B1조

목차

프로젝트 개요

- ✓ 분석 배경
- ✓ 잠재 인자

분석 계획

- ✓ 분석 환경

데이터 구성

- ✓ 데이터 소개
- ✓ 전처리 및 파생변수
- ✓ 탐색적 기법

데이터 분석

- ✓ 분석 방법
- ✓ 모델 종합 평가

결론

- ✓ 핵심 최종 인자
- ✓ 문제점 및 개선점

분석 배경

- ✓ 선박 제조에 사용되는 후판 제품의 Scale 불량 급증이라는 이슈 발생
- ✓ 원인 분석 결과, scratch, 압연흠 등으로 인해 다양한 불량이 발생
- ✓ 그 중 특히 압연공정에서 급증한 것을 확인

불량의 근본 원인을 찾고 불량 예측 및 개선 기회 도출

03 데이터 구성

데이터 소개

plate_no	Plate번호
rolling_date	열연작업시각
scale	Scale(산화철) 불량
spec_long	제품 규격
spec_country	제품 규격 기준국
steel_kind	강종
pt_thick	Plate(후판) 지시두께(mm)
pt_width	Plate(후판) 지시폭(mm)
pt_length	Plate(후판) 지시길이(mm)
hsb	HSB(Hot Scale Braker)적용여부
fur_no	가열로 호기
fur_input_row	가열로 장입열
fur_heat_temp	가열로 가열대 소재온도(℃)
fur_heat_time	가열로 가열대 재로시간(분)
fur_soak_temp	가열로 균열대 소재온도(℃)
fur_soak_time	가열로 균열대 재로시간(분)
fur_total_time	가열로 총 재로시간(분)
fur_ex_temp	가열로 추출온도((℃),계산치)
rolling_method	압연방법
rolling_temp	압연온도(℃)
descaling_count	압연Descaling 횟수
work_group	작업조

목표변수

✓ scale : Scale(산화철) 불량 여부

설명변수

✓ spec_long : 제품 규격

✓ spec_country : 제품 규격 기준국으로 규격은 국가별로 상이

✓ steel_kind : 강종으로 C(탄소강)과 T(티타늄강)으로 구분

✓ hsb : Hot Scale Braker 적용 여부

✓ fur_input_row : 가열로 장입열

✓ rolling_method : 압연방법으로 TMCP(온도제어)와 CR(제어압연)으로 구분

✓ work_groiup : 작업조를 말하는 변수로 4조 2교대

* 카이제곱 검정 결과 유의한 변수 기준으로 설명변수들 중 일부만 작성

잠재 인자

Scale 발생 없음 ↔ 발생	가열로 가열대 온도 저 ↔ 고	가열로 균열대 온도 저 ↔ 고	가열로 추출 온도 저 ↔ 고
Hot Scale Breaker 적용 ↔ 미적용	사상 압연 온도 저 ↔ 고	압연간 Descaling온도 증가 ↔ 감소	판두께 후↔ 박

분석환경

분석 라이브러리

- ✓ Numpy
- ✓ pandas
- ✓ matplotlib
- ✓ seaborn

개발 TOOL

- ✓ Anaconda
- ✓ Jupyter notebook

탐색적 기법

- ✓ 카이제곱 검정
- ✓ 로지스틱 회귀분석

모델링 기법

- ✓ 의사결정트리
- ✓ 랜덤 포레스트
- ✓ XGBoost

데이터 전처리

이상치 제거

rolling_temp=0 인 값

878,	870,	881,	869,	820,
860,	836,	832,	841,	933,
856,	863,	0,	864,	845,
853,	851,	840,	846,	834,
911,	935,	915,	923,	913,

'rolling_temp'의 경우 압연 온도는
0도가 될 수 없어 이상치로 판단하여 제거

결측치 확인

```
plate_no          0
rolling_date      0
scale             0
spec_long         0
spec_country      0
steel_kind        0
pt_thick          0
pt_width          0
pt_length         0
hsb               0
fur_no            0
fur_input_row     0
fur_heat_temp     0
fur_heat_time     0
fur_soak_temp     0
fur_soak_time     0
fur_total_time    0
fur_ex_temp       0
rolling_method    0
rolling_temp      0
descaling_count   0
work_group        0
dtype: int64
```

결측치 존재 안함

제거한 행

fur_ex_temp : 가열로 추출온도 겹치는 column이므로 제거

plate_no : plate 번호는 수율에 영향을 주지 못한다고 판단하여 제거

rolling_date : datetime, time 생성 했으므로 기존 시간데이터 제거

time : hour와 겹치는 행이라 제거

spec_long : 데이터 부족으로 유의미한 분석 어렵다 판단해 제거

파생 변수

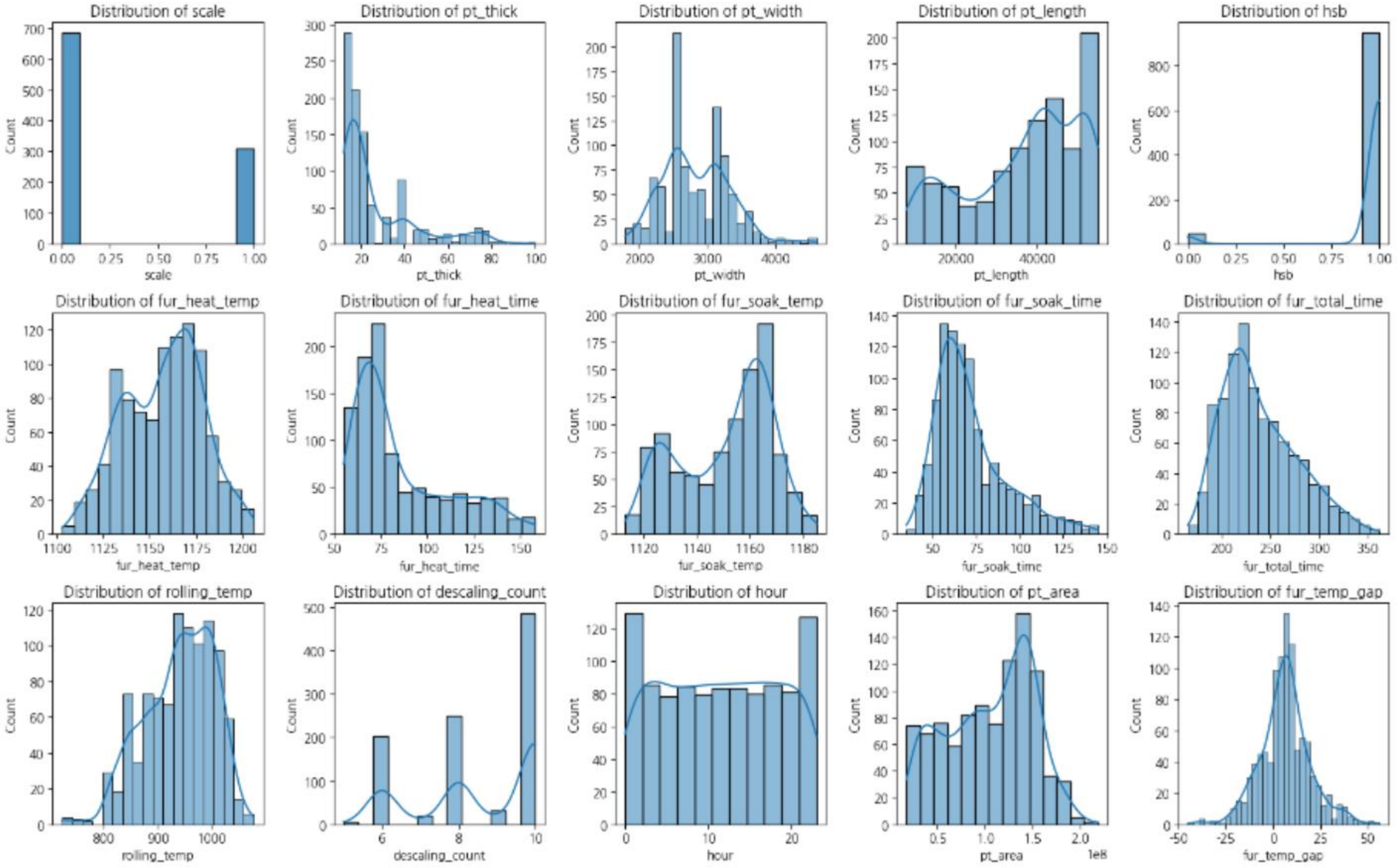
pt_area	후판면적 width * length	큰 면적을 가진 후판일수록 scale발생 가능성이 높다고 판단
rolling_temp_>900	압연온도에 따른 구분 900초과면 1, 이하면 0	900도 이상에서 산화가 기하급수적으로 더 잘 일어난다는 특성을 고려함
fur_temp_gap	가열대와 균열대 온도차	가열대와 균열대 온도차는 재료의 열적응과 열확산 특성을 나타낼 수 있음
hour	rolling_date에서 datetime을 사용해 작업 시간대 추출	작업하는 시간대에 따라 작업자의 피로도와 성능이 변할 수 있음 특히 긴 근무 시간이나 특정 시간대의 피로는 불량 발생률을 높일 수 있다고 판단

03 데이터 구성

탐색적 기법

히스토그램

데이터의 전반적인 분포 파악 및
경향성 확인

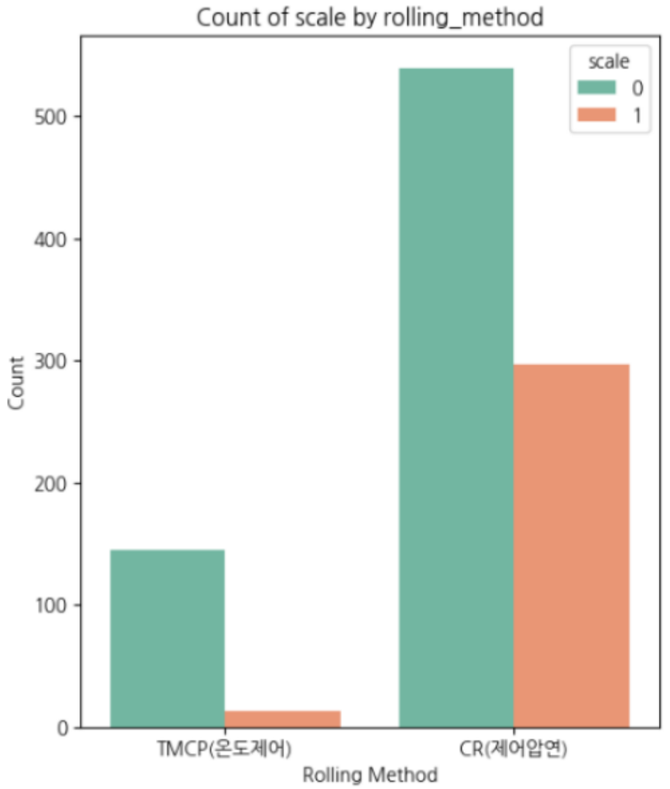


* 노란색 박스는 유의미해 보이는 변수들

03 데이터 구성

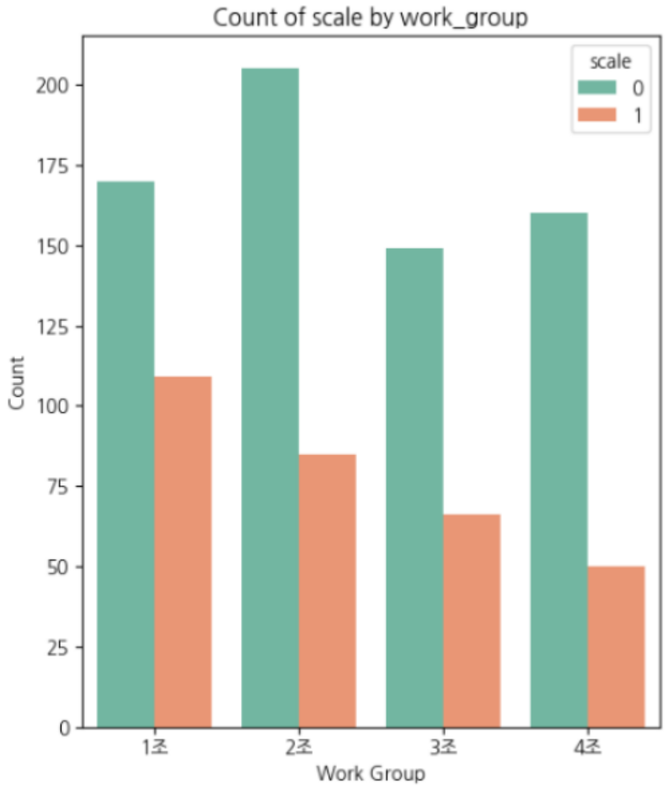
탐색적 기법

barplot



제어압연의 경우 온도제어보다 양품의 비율 ↑

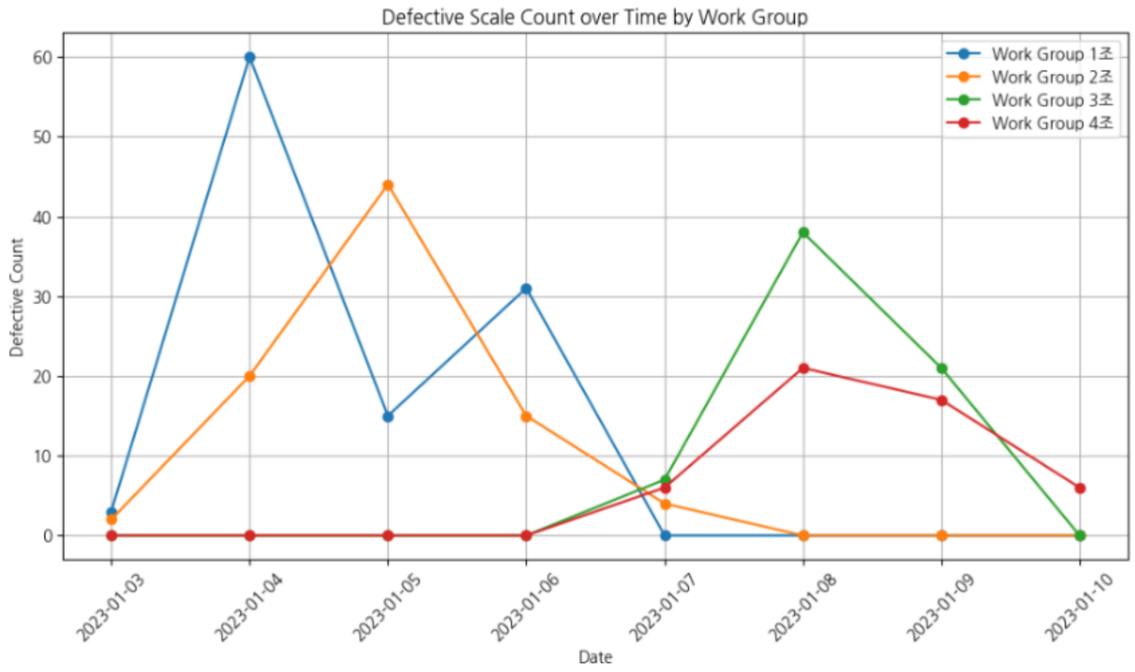
즉, 온도제어의 경우 제어압연보다 불량률 ↑



작업조의 경우 1조가 불량률 가장 ↑

즉, 작업시간대가 불량률에 영향을 미칠수도 있음

Lineplot



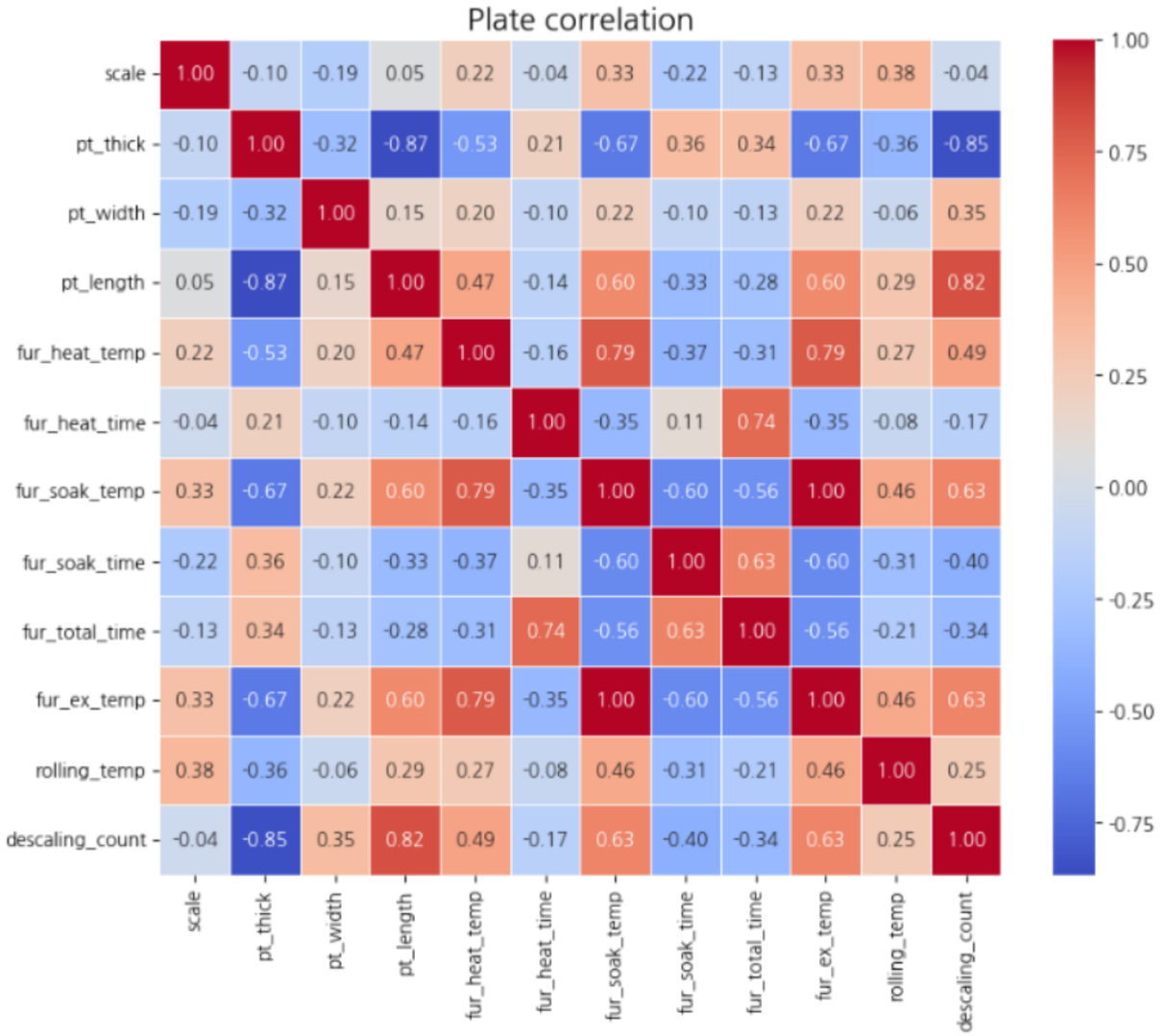
불량율이 0인곳을 보면 해당일자에 근무를 안했음 파악 가능

또한 1,2 조가 3,4조에 비해 불량률이 높으며

4조가 가장 적은 불량률을 보임

03 데이터 구성

탐색적 기법



상관관계 분석

- ✓ 1.0인 fur_total_time변수는 데이터내 중복된 열이라 제거
- ✓ pt_thick와 pt_length의 상관계수는 -0.87로 음의 상관관계를 보임
- ✓ pt_thick와 descaling_count의 상관계수는 -0.85로 음의 상관관계를 보임
- ✓ pt_length와 descaling_count의 상관계수는 0.82로 양의 상관관계를 보임

탐색적 기법

카이제곱 검정 : 범주형과 범주형 변수간의 검정

```
Chi-square test for spec_long:
Chi-square test statistic: 235.94685749378158
P-value: 3.1133164751445 94e-21
유의수준 0.05에서 귀무가설 기각: 'scale'과 spec_long 사이에는 유의한 관련성이 있다.
```

```
Chi-square test for spec_country:
Chi-square test statistic: 69.80075036426382
P-value: 4.4922555761885164e-13
유의수준 0.05에서 귀무가설 기각: 'scale'과 spec_country 사이에는 유의한 관련성이 있다.
```

```
Chi-square test for steel_kind:
Chi-square test statistic: 76.25774182995244
P-value: 2.489547428454086e-18
유의수준 0.05에서 귀무가설 기각: 'scale'과 steel_kind 사이에는 유의한 관련성이 있다.
```

```
Chi-square test for hsb:
Chi-square test statistic: 105.51048606504317
P-value: 9.439705302426995e-25
유의수준 0.05에서 귀무가설 기각: 'scale'과 hsb 사이에는 유의한 관련성이 있다.
```

```
Chi-square test for fur_no:
Chi-square test statistic: 3.1186222255276252
P-value: 0.2102808811312071
유의수준 0.05에서 귀무가설 채택: 'scale'과 fur_no 사이에는 유의한 관련성이 없다.
```

```
Chi-square test for fur_input_row:
Chi-square test statistic: 0.9203672191669445
P-value: 0.3373785709791819
유의수준 0.05에서 귀무가설 채택: 'scale'과 fur_input_row 사이에는 유의한 관련성이 없다.
```

```
Chi-square test for rolling_method:
Chi-square test statistic: 44.88003167017692
P-value: 2.0948325380804966e-11
유의수준 0.05에서 귀무가설 기각: 'scale'과 rolling_method 사이에는 유의한 관련성이 있다.
```

```
Chi-square test for work_group:
Chi-square test statistic: 13.900577240005461
P-value: 0.003043655675169395
유의수준 0.05에서 귀무가설 기각: 'scale'과 work_group 사이에는 유의한 관련성이 있다.
```

그 결과 총 8개중 6개의 유의미한 변수 도출

spec_long, spec_country, steel_kind, hsb, fur_input_row, rolling_method, work_group

탐색적 기법

로지스틱 회귀분석 : 범주형과 연속형 변수간의 검정

```
Optimization terminated successfully.
         Current function value: 0.412833
         Iterations 7
                           Logit Regression Results
==============================================================================
Dep. Variable:                  scale   No. Observations:                  994
Model:                          Logit   Df Residuals:                      988
Method:                           MLE   Df Model:                            5
Date:                Mon, 11 Mar 2024   Pseudo R-squ.:                  0.3348
Time:                        13:12:10   Log-Likelihood:                -410.36
converged:                       True   LL-Null:                       -616.87
Covariance Type:            nonrobust   LLR p-value:                 4.630e-87
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -92.4601     10.315     -8.964      0.000    -112.676     -72.244
pt_width       -0.0006      0.000     -2.968      0.003      -0.001      -0.000
fur_heat_time   0.0104      0.004      2.641      0.008       0.003       0.018
fur_soak_temp   0.0683      0.010      6.890      0.000       0.049       0.088
rolling_temp    0.0201      0.002      9.218      0.000       0.016       0.024
descaling_count -0.6427     0.085     -7.603      0.000      -0.808      -0.477
==============================================================================
```

로지스틱 회귀분석 결과, 다음의 변수들은 scale에 유의한 영향을 미침

pt_width, fur_heat_time, fur_soak_temp, rolling_temp, descaling_count

	Variable	VIF
0	fur_heat_time	7.50341
1	descaling_count	7.50341

로지스틱 회귀분석 결과, 유의한 변수들중에
다중공선성을 보이지 않는 변수는 다음과 같다

fur_heat_time, descaling_count

04 데이터 분석

분석 방법 : 의사결정트리

Train / Test set

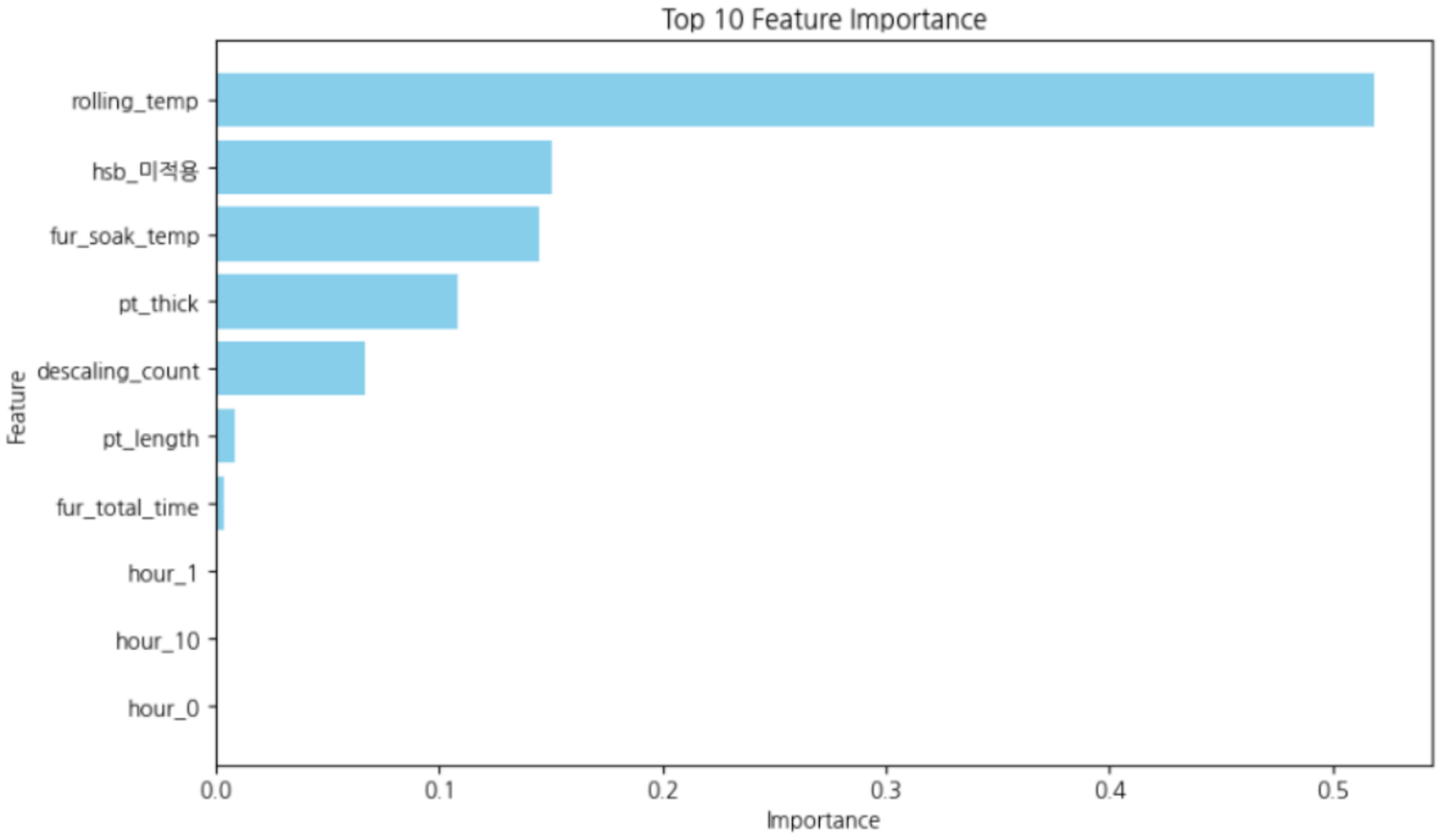
```
분할 전 설명변수 데이터 : (994, 59)
분할 후 설명변수 데이터 :Train (695, 59)   Test (299, 59)
```

→ Train / Test 는 0.3의 비율로 분리

모델 설명력

```
Accuracy on training set: 1.000
Accuracy on test set: 0.993
```

→ train data의 설명력은 과대적합의 위험이 있으나
test data 모델 설명력은 0.993으로 매우 높음

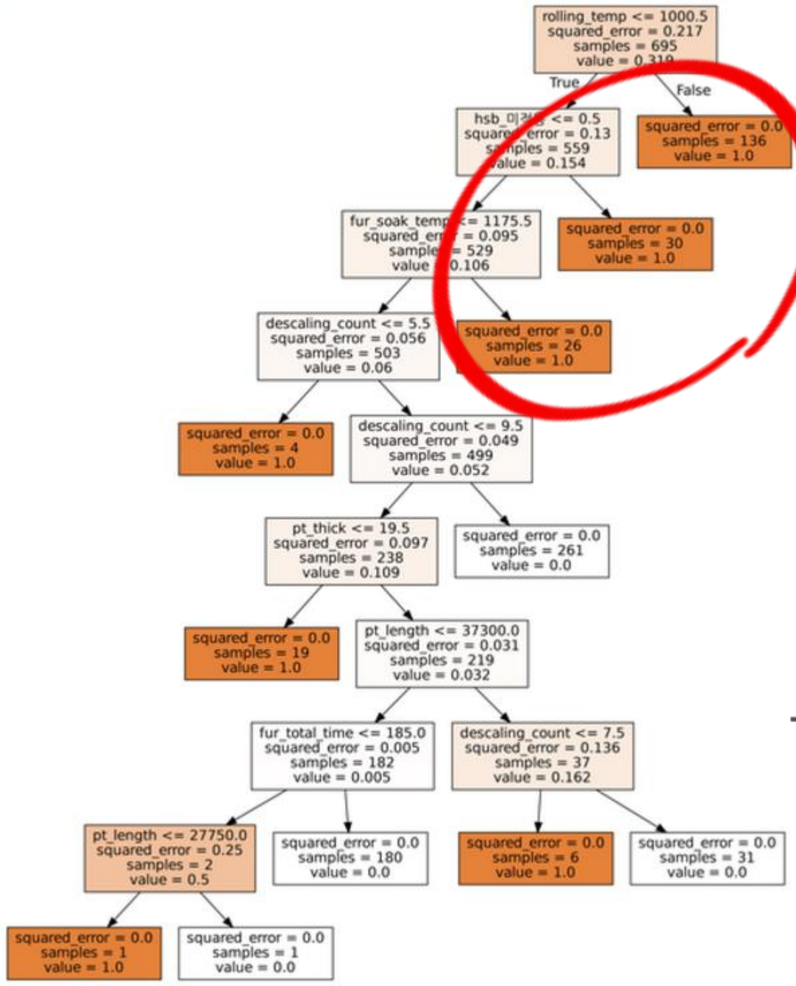


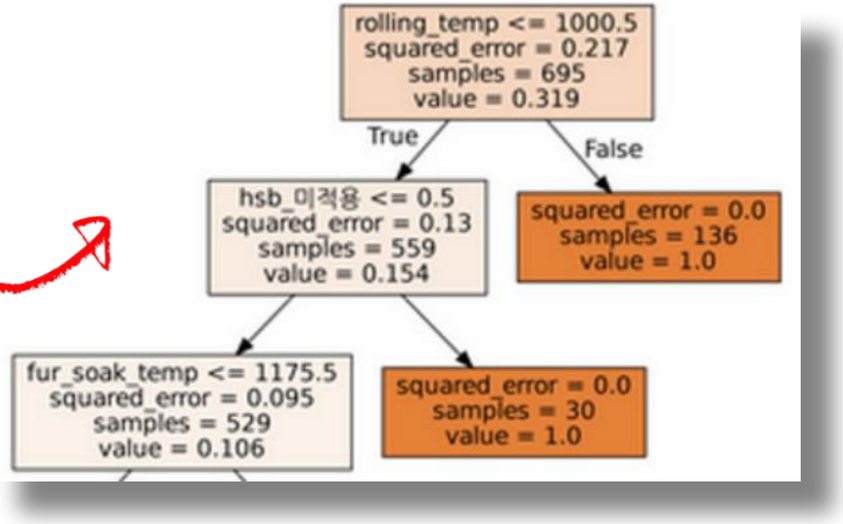
> 의사결정트리의 분석결과, rolling_temp의 설명변수 중요도가 가장 높음
> 즉, 후판공정 scale발생에 있어서 가장 중요한 역할을 수행함
> 그 뒤로는 hsb_미적용, fur_soak_temp가 2, 3위를 차지함

* 모델 설명력 기준은 0.8 이상은 좋은 성능의 모델로 판단

04 데이터 분석

분석 방법 : 의사결정트리





의사결정트리의 분석 결과 rolling_temp를 기준으로

1000.5도 이상이 되었을 때 불량이 발생했음을 알 수 있음

→ 따라서 1000도 이하로 낮춤으로써 불량률을 줄일 수 있을것으로 판단됨

04 데이터 분석

분석 방법 : 랜덤포레스트

Train / Test set

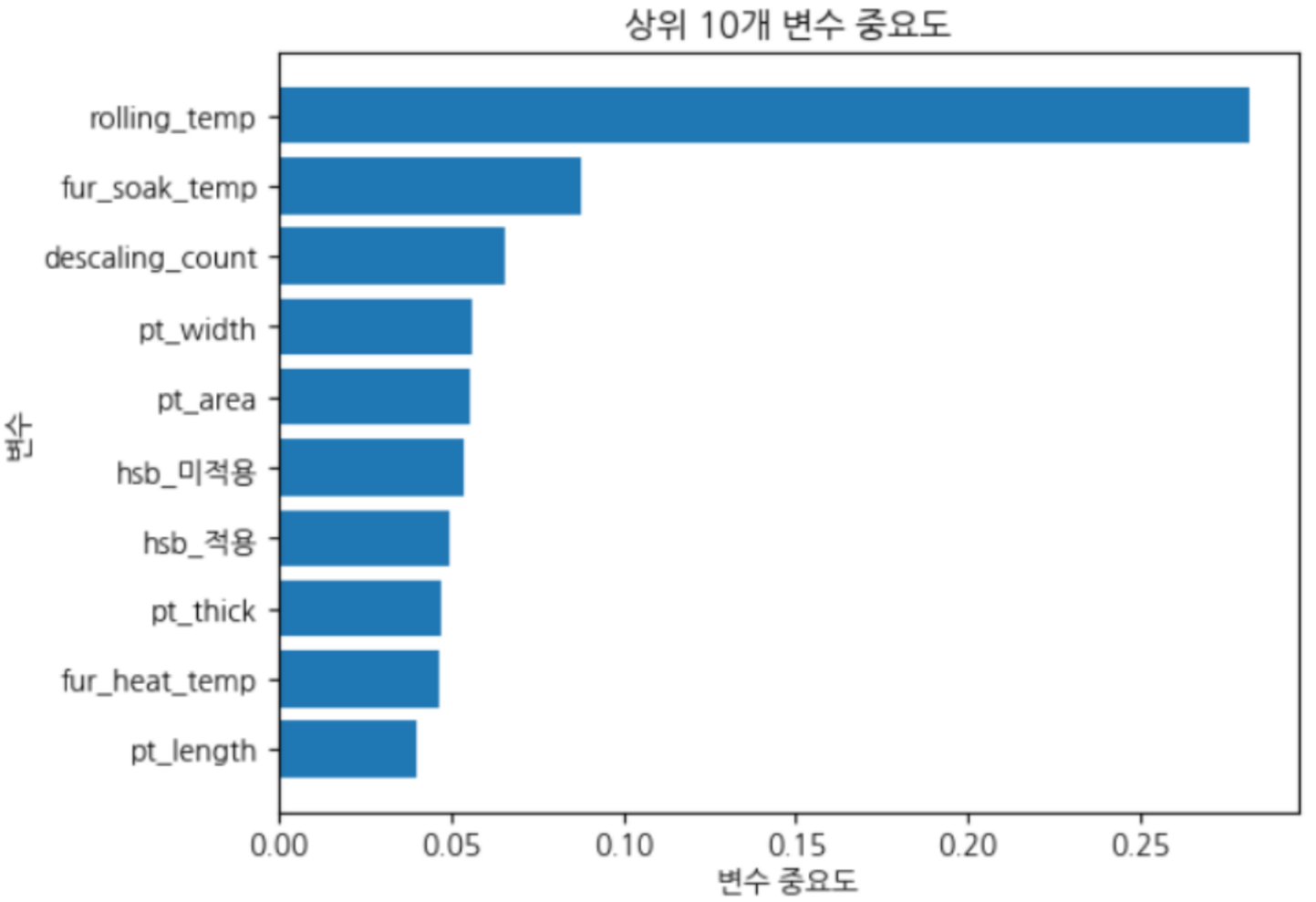
```
분할 전 설명변수 데이터 : (994, 59)
분할 후 설명변수 데이터 :Train (695, 59)   Test (299, 59)
```

→ Train / Test 는 0.3의 비율로 분리

모델 설명력

```
Score on training set: 0.965
Score on test set: 0.960
```

→ test data 모델 설명력은 0.960으로 상당히 높은편



랜덤포레스의 분석결과, rolling_temp의 설명변수 중요도가 가장 높음
즉, 후판공정 scale발생에 있어서 가장 중요한 역할을 수행함
그 뒤로는 fur_soak_temp, descaling_count가 2, 3위를 차지함

* 모델 설명력 기준은 0.8 이상은 좋은 성능의 모델로 판단

04 데이터 분석

분석 방법 : XGBoost

Train / Test set

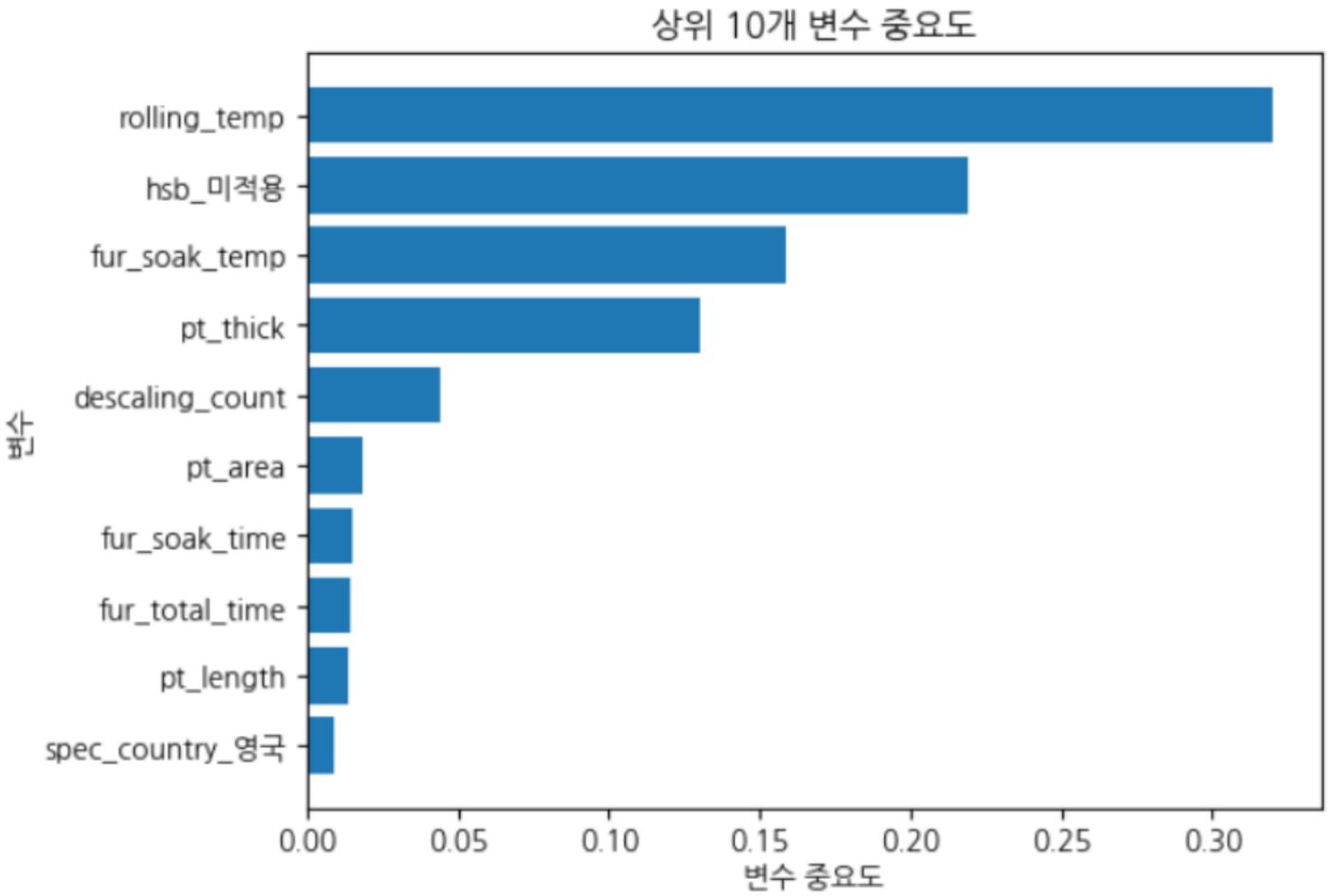
```
분할 전 설명변수 데이터 : (994, 59)
분할 후 설명변수 데이터 :Train (695, 59)   Test (299, 59)
```

→ Train / Test 는 0.3의 비율로 분리

모델 설명력

```
Score on training set: 1.000
Score on test set: 0.990
```

→ train data의 설명력은 과대적합의 위험이 있으나
test data 모델 설명력은 0.990으로 매우 높음



XGBoost의 분석결과, rolling_temp의 설명변수 중요도가 가장 높음
즉, 후판공정 scale발생에 있어서 가장 중요한 역할을 수행함
그 뒤로는 hsb_미적용, fur_soak_temp가 2, 3위를 차지함

* 모델 설명력 기준은 0.8 이상은 좋은 성능의 모델로 판단

모델 종합 평가

Accuracy

의사결정트리

```
Accuracy on training set: 1.000
Accuracy on test set: 0.993
```

랜덤포레스트

```
Score on training set: 0.965
Score on test set: 0.960
```

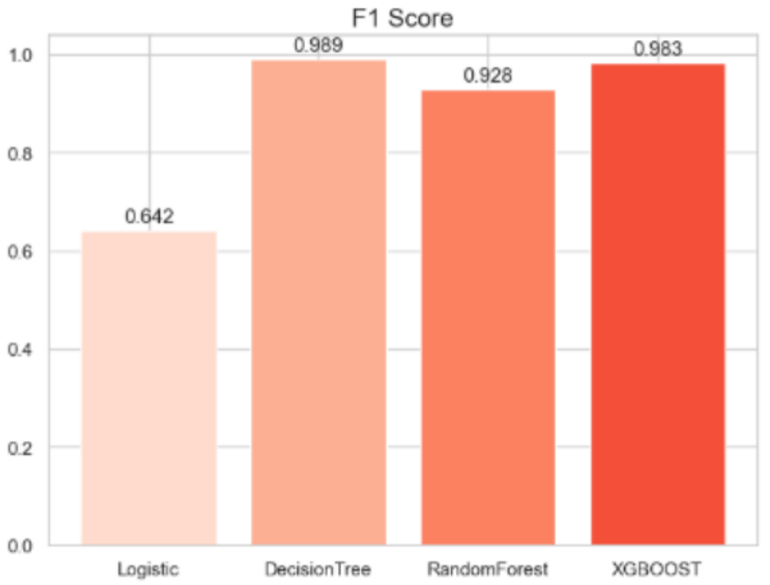
XGBoost

```
Score on training set: 1.000
Score on test set: 0.990
```

train data의 설명력은 과대적합의 위험이 있으나
test data 모델 설명력 또한 높아 일반화가 가능하다고 판단

→ 정확도는 0.993인 의사결정트리가 가장 높음

F1 score



로지스틱 회귀분석은 f1 score의 설명력이 너무 낮아 제외

→ F1 score는 0.989인 의사결정트리가 가장 높음

> 모델의 성능과 일반화 능력이
> 가장 뛰어난 의사결정트리 선정

05 결론

핵심 최종 인자

rolling_temp → '압연온도' 가 불량의 주요 원인으로 예측됨

고온에서의 압연 공정은 강판 표면의 화학적 반응을 촉진하기 때문에 철과 산소가 반응하여 산화철 스케일이 형성될 가능성 증가

압연 공정 라인에 **온도 모니터링 시스템을 도입**하여 실시간으로 압연 온도 추적, 안정적인 온도 범위 내에서 **자동 제어 시스템 도입**

fur_soak_temp → '가열로 균열대 소재 온도' 가 불량의 주요 원인으로 예측됨

가열로 균열대 고온 처리 과정은 증류수와 철강의 화학적 반응을 촉진시켜 저온 처리 과정보다 두꺼운 산화철 스케일 형성함

균열대 소재 온도가 비정상적으로 오를 때, 이상 감지를 통해 가열로의 온도 조건을 조절할 필요가 있음

hsb → 'HSB 적용여부' 가 불량의 주요 원인으로 예측됨

열연 공정 과정 중 생기는 철강 표면에서의 결함은 초기 결함의 부피에 비례하여 증가하고 이는 품질저하의 주요 원인임

열연 공정 전 hsb를 통한 결함제거가 중요한 공정 요소임

05 결론

문제점 및 개선점

압연 온도 제어 → 압연 온도를 1000도 이하로 낮추는 것이 불량품을 줄이는데 효과적이다.

문제점

- ✓ 기존의 압연공정은 1000도 이상에서 산화반응이 기하급수적으로 증가하는 임계온도에 도달한 것으로 예측됨
- ✓ 약 700도에서 1000도 사이로 압연 공정의 온도 조건을 맞춰서 운전해야 양품을 생산하는데 유리함

개선점

- ✓ 최적의 공정 온도 조건을 맞추기 위해 라인에 온도 센서나 자동 제어 장치와 같은 공정 제어 시스템 도입
- ✓ 하지만 생산성 향상 대비 설비 설치비용, 설비 유지비용을 고려하여 공정의 경제성을 평가해야 함
- ✓ 약 700도에서 1000도 사이의 온도에서 공정을 운전할 때, 열 에너지 소모 비용 등 경제성을 고려하여 최적의 온도에서 운전해야 함

그 외 고려해야 할 점

가열로 재로 시간

가열로 재로 시간 ↓ , 불량품 생산 가능성이 ↑

가열로 재로 시간 ↓ , 철강 생산성이 향상하기 때문에 최적의 재로 시간을 고려해야 함

압연 descaling 횟수

압연 descaling 횟수가↓, 불량품 생산 가능성이 ↑

압연 descaling 횟수가 ↓, 설비 유지 비용이 절약되기 때문에 최적의 압연 descaling 횟수를 고려해야 함

표면적

표면적이 ↑, 산화철의 형성이 두꺼워지기 때문에 불량품 생산 가능성이 증가함

표면적이↑, 단위 면적당 생산 단가↑ , 따라서 고객사의 주문에 맞춰 최적의 표면적으로 생산해야 함

posco

감사합니다

청년 AI·Big Data 아카데미 25기

B1조