The background is a dark industrial scene at night, featuring a large metal structure with a staircase on the right and bright overhead lights. A large white circle on the left contains a close-up of a blue industrial machine part. A solid orange circle is positioned at the bottom center, overlapping the white circle and the background.

후판공정 불량 예측모델 개발 및 핵심요소 파악

C1 권구택

목차

01 프로젝트 정의 및 배경
후판 공정 불량 개선

02 데이터 현황 파악
설명 변수 내용 파악 및 결측치, 이상치 확인

03 데이터 EDA
그래프 기반 분포 확인 및 통계 검정을 통한 분석

04 모델링 및 모델 성능 개선
트리 기반 모델 모델링

05 개선안 도출
모델 해석 및 EDA 기반 개선안 도출



01

프로젝트 정의 및 배경

1. 프로젝트 배경
2. 프로젝트 목표
3. 프로젝트 개요
4. 프로젝트 내용

프로젝트 정의 및 배경



프로젝트 배경

- 최근 들어 선박 제조에 주로 사용되는 후판 제품에 'Scale 불량 급증'이라는 이슈가 발생
- 원인 분석 결과, 압연흙, Scratch 등 다양한 불량들이 발생하고 있으나 특히 압연 공정에서 불량이 급증한 것을 확인

프로젝트 목표

- 수집된 데이터 기반 불량 근본 원인을 찾고 불량 예측 및 개선 기회를 도출하고자 함

프로젝트 개요

- 불량 발생에 영향을 미치는 요인 기반 영향 인자 선정 및 데이터 기반 불량 예측 모델 개발

프로젝트 내용

- 다양한 기법 기반 후판 공정 불량률에 영향을 주는 잠재 인자 도출
- 도출된 설명 변수와 사전 검토한 잠재 요인과의 일치, 방향성 등을 확인하고 정리
- 예측 모델 개발 및 해당 모델 특징 기반 성능 개선 방안 도출



02

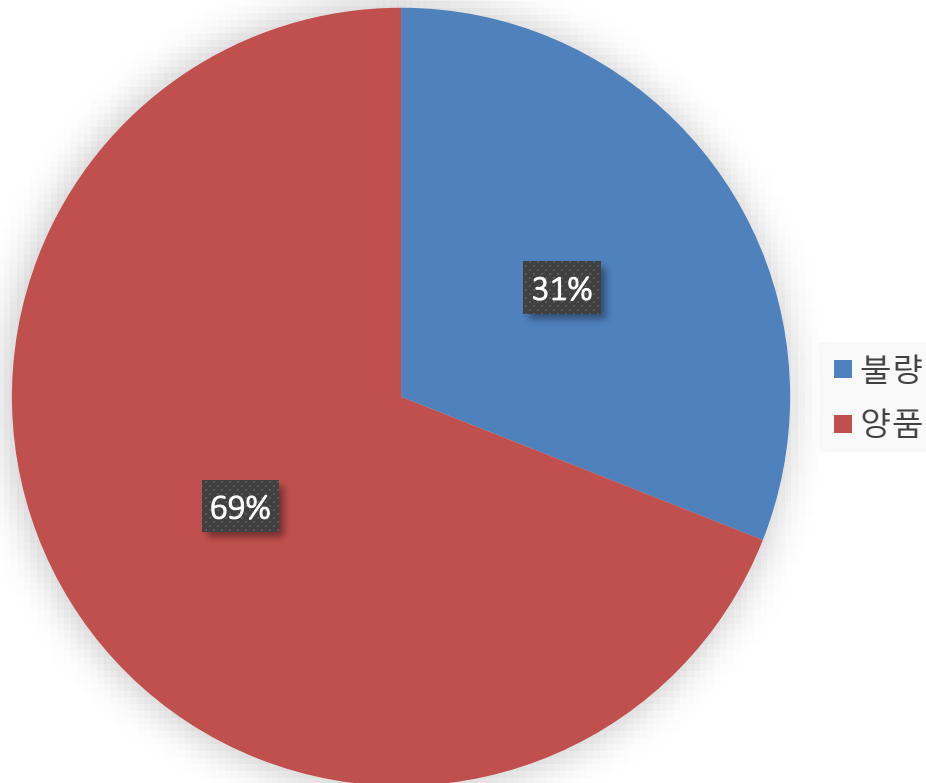
데이터 현황 파악

1. 변수 설명
2. 결측치 확인
3. 이상치 확인
4. VF 탐색

변수 설명

변수명	역할	Dtype	변수 설명	척도
plate_no	ID	범주형	Plate번호	Nominal
rolling_date	날짜	연속형	열연작업시각	Datetime
scale	목표변수	범주형	Scale(산화철) 불량	Binary
spec_long	설명변수	범주형	제품 규격	Nominal
spec_country	설명변수	범주형	제품 규격 기준국	Nominal
steel_kind	설명변수	범주형	강종	Nominal
pt_thick	설명변수	연속형	Plate(후판) 지시두께(mm)	Interval
pt_width	설명변수	연속형	Plate(후판) 지시폭(mm)	Interval
pt_length	설명변수	연속형	Plate(후판) 지시길이(mm)	Interval
hsb	설명변수	범주형	HSB(Hot Scale Braker)적용여부	Binary
fur_no	설명변수	범주형	가열로 호기	Nominal
fur_input_row	설명변수	범주형	가열로 장입열	Nominal
fur_heat_temp	설명변수	연속형	가열로 가열대 소재온도(°C)	Interval
fur_heat_time	설명변수	연속형	가열로 가열대 재로시간(분)	Interval
fur_soak_temp	설명변수	연속형	가열로 균열대 소재온도(°C)	Interval
fur_soak_time	설명변수	연속형	가열로 균열대 재로시간(분)	Interval
fur_total_time	설명변수	연속형	가열로 총 재로시간(분)	Interval
fur_ex_temp	설명변수	연속형	가열로 추출온도((°C),계산치)	Interval
rolling_method	설명변수	연속형	압연방법	Interval
rolling_temp	설명변수	범주형	압연온도(°C)	Binary
descaling_count	설명변수	연속형	압연Descaling 횟수	Interval
work_group	설명변수	범주형	작업조	Nominal

데이터 분포 확인 - 불량률



- 불량률이 31%로 매우 높다.
- 개선안을 도출하지 않으면 후판 시장에서 경쟁력을 잃을 수 있다.
- 국제 표준인 ISO 9001 등의 인증을 받기 위해서는 특정 불량률 기준을 충족해야 한다.

결측치 및 이상치 확인



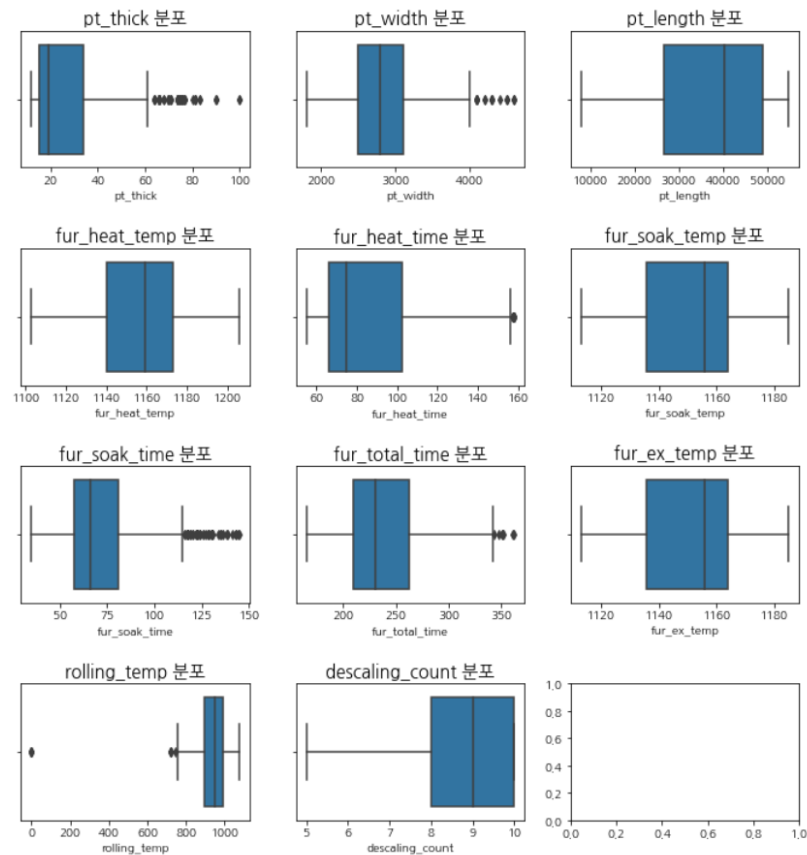
결측치

해당 데이터에서는 결측치가 없음

이상치

그래프 및 기술적 통계 확인 결과, rolling_temp에서 이상치를 발견하고 해당 값을 평균으로 대체

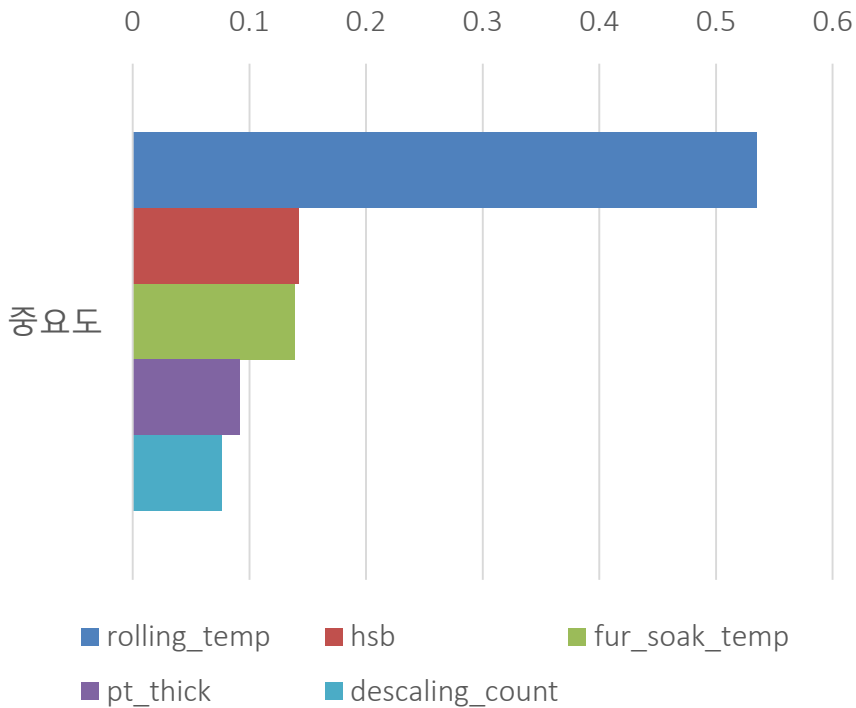
판단근거 : 시장 조사 결과, 회사별 차이가 있지만 압연 온도의 적정 범위는 1000~1100도 사이기 때문에 rolling_temp 컬럼에서 0은 이상치로 판단



Vital Few 탐색



Vital Few (핵심원인변수)



Vital Few 탐색 결과

- Rolling_temp, hsb, fur_soak_temp, pt_thick, descaling_count가 영향을 많이 주는 변수다.
- 공정 과정에서 추가 공정 횟수 및 공정 온도가 큰 영향을 끼치는걸 알 수 있다.



03

데이터 EDA

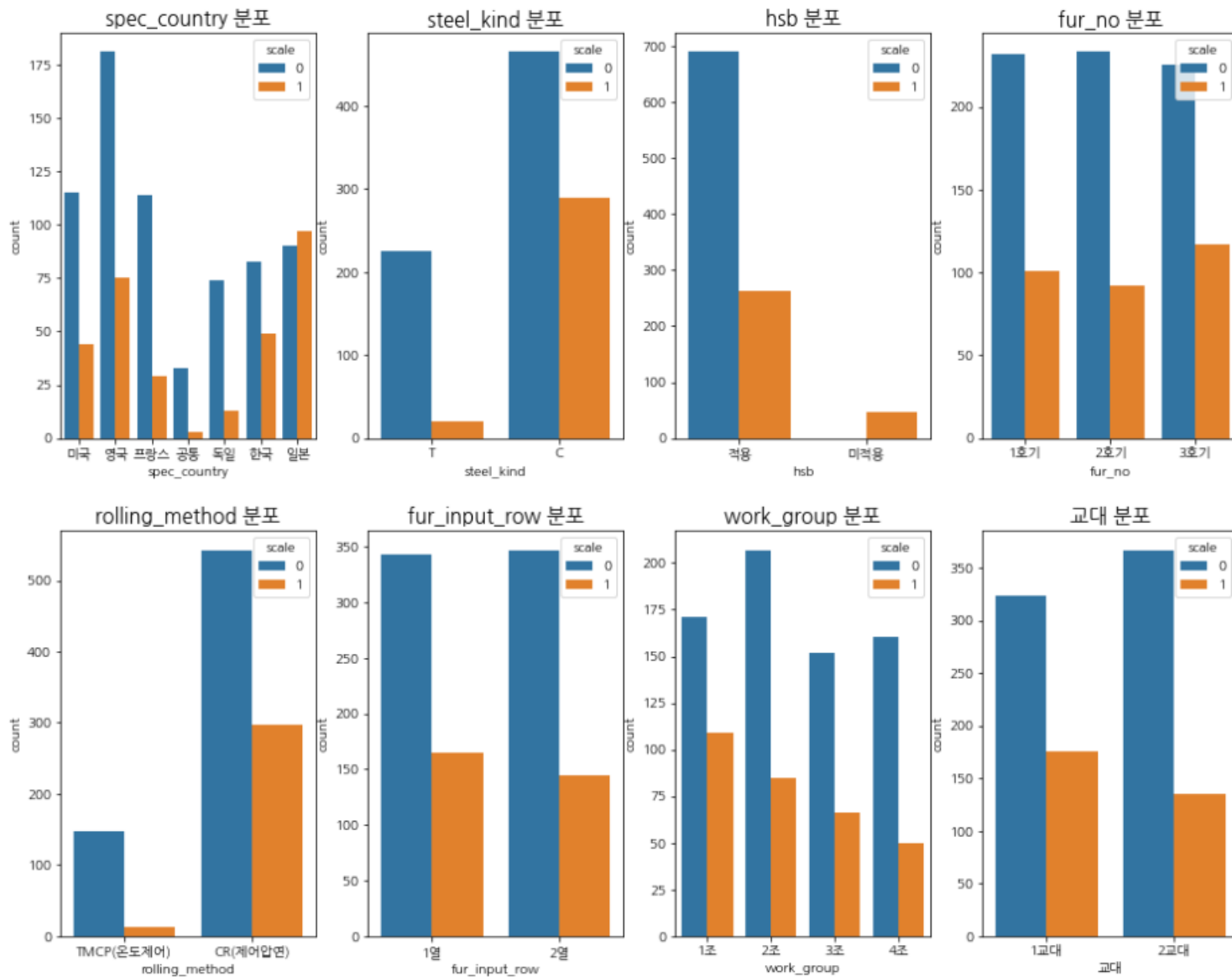
1. 파생 변수
2. 설명 변수 분포 확인
3. 변수 상관계수 확인
4. 제거 예정 변수 선정

파생 변수

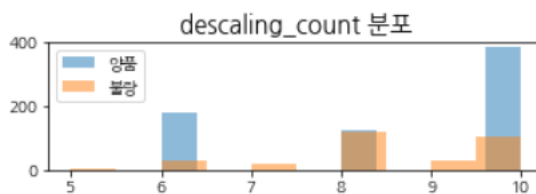
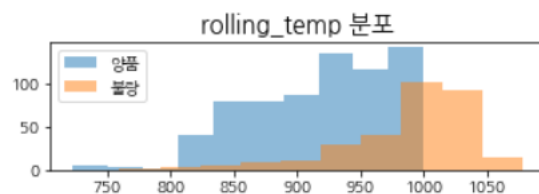
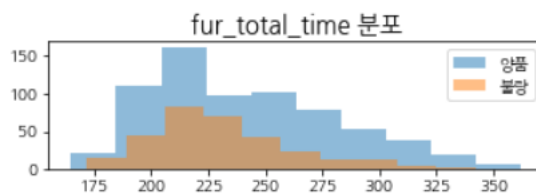
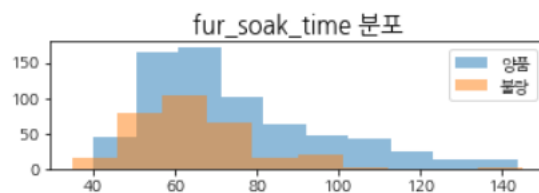
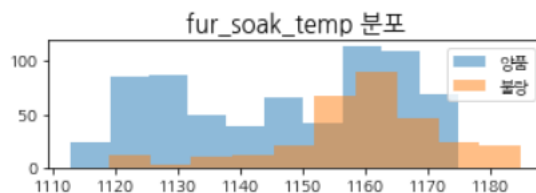
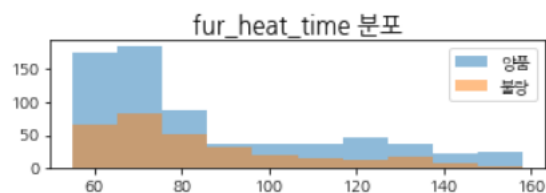
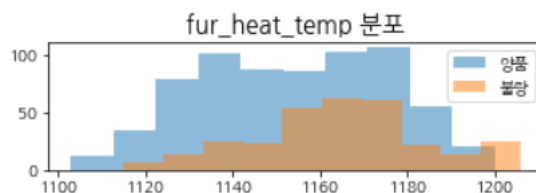
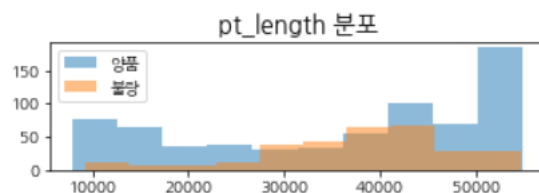
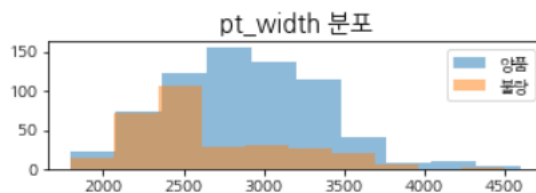
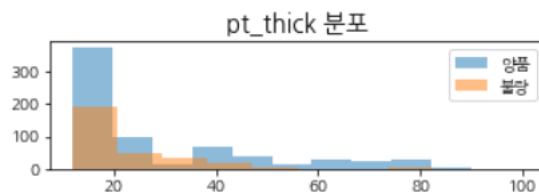


파생 변수 명	파생 근거	적용 여부
Fur_remain_time	<ul style="list-style-type: none"> - fur_total_time은 가열로 총 재로 시간인데 가열대 재로시간 및 균열대 재로시간의 합과 차이가 있어서 온도 냉각 시간대가 있다고 판단하여 파생 - 그래프 분석 결과 해당 변수와 불량률에 유의미한 차이가 없다고 판단 	X
교대	<ul style="list-style-type: none"> - 각 work_group마다 정해진 시간이 있어서 해당 교대 근무조에 따른 성과 차이가 있지 않을까라는 가정 하에 파생 - 교대 시간에 따른 불량률에 유의미한 차이가 있어서 선정 	O
호열	<ul style="list-style-type: none"> - fur_input_row 컬럼과 fur_no 컬럼에서 파생되는 변수로 각 호기에서 각 열마다의 데이터가 다르다고 생각해서 결합 필요를 느낌 - 검정 시에 해당 컬럼에 따른 불량률에 유의미한 차이가 없어서 선정하지 않음 	X
가열대 기울기	<ul style="list-style-type: none"> - 가열대 내에서 가열되는 속도가 얼마나 빠르지에 따라서 공정에 있어서 유의미한 차이가 있는지 판단하기 위해 파생 - 검정 시에 해당 컬럼이 유의미하지 않기 때문에 선정하지 않음 	x

변수 분포 확인



변수 분포 확인



변수 분포 확인



- HSB(Hot Scale Breaker)을 적용하지 않으면 100% 불량 발생한다. 즉, HSB는 공정에서 무조건 적용해야 한다는 걸 알 수 있다.
- 국가 규격별 불량률 분포 확인 결과, 한국 규격에서 발생하는 불량률이 높으며, 일본은 불량률이 양품률보다 높다.
 - 일본 기준 규격 데이터만 따로 추출해서 분석 진행
- 강종이 탄소인 제품들의 불량률이 매우 높다.
 - 탄소 강종 데이터들만 따로 추출해서 분석 진행
- Rolling-method가 제어 압연일 경우, 높은 불량률을 보인다.
 - CR(제어 압연) 방식으로 압연을 진행하는 데이터만 추출해서 분석 진행

변수 분포 확인



- work_group과 교대 컬럼에서도 다른 불량률을 가지고 있다고 보인다.
 - 통계 기법을 기반으로 유의미한 차이가 있는지 검정 진행
- 기타 범주형 설명 변수들은 그래프 상 불량률에 유의미한 차이가 없다고 판단했다.
- pt_thick 컬럼에서 두께가 얇을 때 비교적 불량률이 발생하는 경우가 많은 걸 알 수 있다.
 - 얇은 두께 공정에 있어서 공정 과정에서 많은 어려움이 있는지 파악해 봐야한다.
- rolling_temp가 1000도를 넘으면 모든 제품이 불량이라는 걸 알 수 있다.
 - 사전조사에서는 1000도에서 1100도가 적정 온도라고 했는데 해당 회사의 공정에서는 1000도 이상이면 100% 불량률이 발생한다.
 - 1000도 이상 공정에서 어떤 문제가 발생해서 그러는지 파악해볼 수 있다.

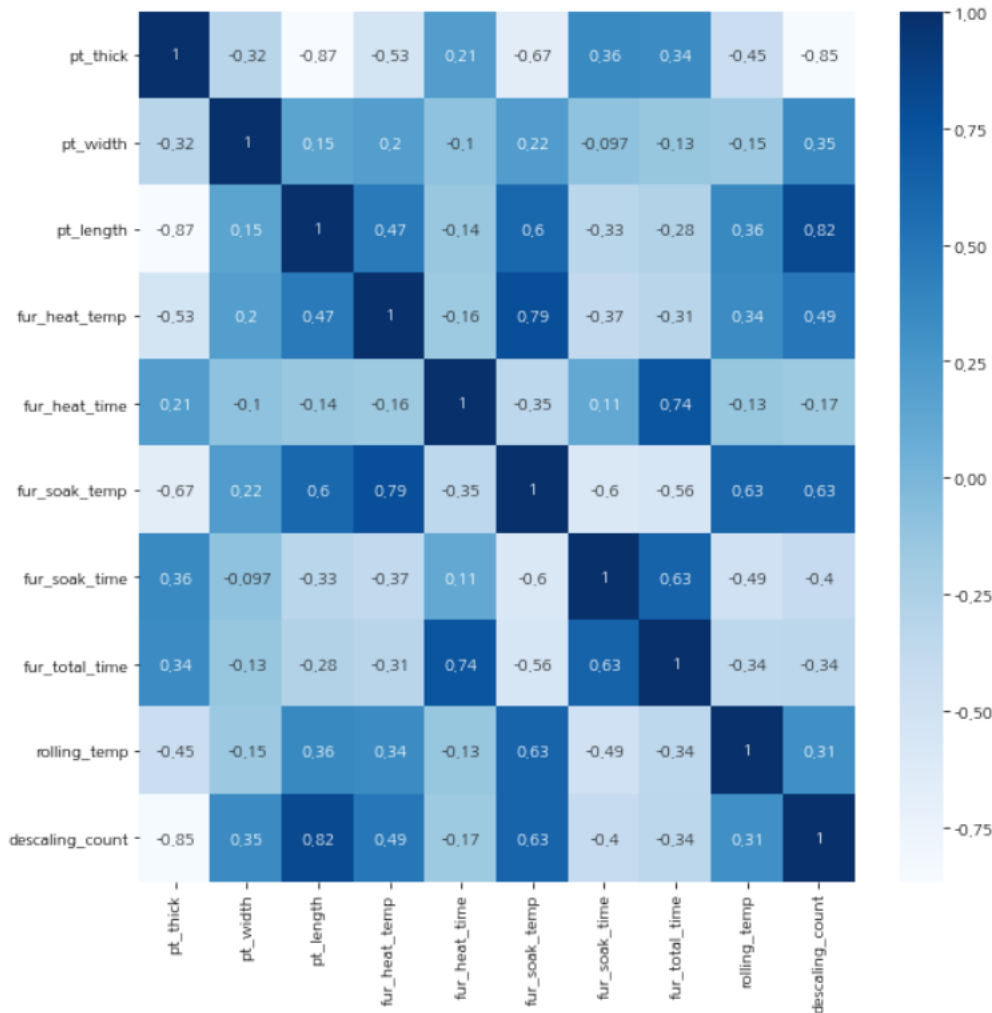
변수별 상세 분석



변수 명	분석 내용
spec_long	<ul style="list-style-type: none"> - 각 규격별로 기준 국가, 내부 함량 등이 다 다르기 때문에 특정 규격의 불량률에 대해서 판단 가능하나 전체 불량률에 대해서 판단하기에는 해당 변수의 설명력이 부족하다고 판단 - 그래프 분석 결과, 100% 양품인 규격과 100% 불량인 규격이 존재 <ul style="list-style-type: none"> - 100% 양품인 제품은 HSB를 적용한 제품이다. - 모든 조가 모든 규격을 담당하지 않는다. - 1호기 2열에서는 100% 불량인 해당 규격들을 공정한 적이 없다.
spec_country	<ul style="list-style-type: none"> - H_0: 범주별 불량률에 유의미한 차이가 없다. - 통계 검정에서 p-value가 유의 수준보다 작기 때문에 귀무가설을 기각하고 대립가설을 채택 - 즉, 해당 변수들은 변수별 불량률에 유의미한 차이가 있다.
steel_kind	
hsb	
work_group	
교대	
rolling_method	
호열	<ul style="list-style-type: none"> - 귀무가설 채택으로 불량률에 유의미한 차이가 없다

pt_thick, pt_width, fur_heat_temp, fur_soak_temp, fur_soak_time, fur_total_time, rolling_temp
-> 해당 컬럼들도 불량률에 유의하다는 것을 알 수 있음

변수 상관관계수 확인



강한 상관관계를 가진 변수들의 조합은 다음과 같다. (상관계수 > 0.5)

- pt_length, fur_soak_tem
- pt_length, descaling_count
- fur_heat_temp, fur_soak_temp
- fur_heat_time, fur_total_time
- fur_soak_temp, pt_length
- fur_soak_temp, fur_heat_temp
- fur_soak_temp, rolling_temp
- fur_soak_temp, descaling_count
- fur_soak_time, fur_total_time
- fur_total_time, fur_heat_time
- fur_total_time, fur_soak_time
- rolling_temp, fur_soak_temp
- descaling_count, pt_length
- descaling_count, fur_soak_temp

제거 예정 변수 선정



제거 예정 변수	제거 근거
plate_no	인덱스 용도로 사용되기 때문에 분석에 필요 없다고 판단해서 삭제
rolling_date	해당 변수에서 필요하다고 판단한 부분은 파생변수로 생성하였고 시계열 분석을 진행하기에는 데이터의 수가 매우 적다고 판단해서 삭제
fur_ex_temp	fur_soak_temp랑 동일하기 때문에 삭제
fur_input_row	호열이란 파생변수를 생성하는데 사용해서 해당 변수를 삭제
fur_no	호열이란 파생변수를 생성하는데 사용해서 해당 변수를 삭제
spec_long	분석 과정에서 해당 변수로 전체적인 불량률을 판단하는데 유의미하지 않다고 판단해서 삭제
pt_length	불량률 판단에 있어서 유의미하지 않다고 검정되어서 삭제
fur_heat_time	불량률 판단에 있어서 유의미하지 않다고 검정되어서 삭제
work_group	교대라는 파생 변수에 사용되었기 때문에 해당 변수는 삭제



04

모델링

1. 앙상블 모델
2. 최고 성능 모델 선정

모델 종류	데이터 종류	Accuracy
랜덤포레스트	train 데이터	1
	test 데이터	0.96
XGB	train 데이터	1
	test 데이터	0.9867
LGBM	train 데이터	1
	test 데이터	1

- 해당 모델 성과는 기본 파라미터를 이용한 모델 기반 성능이다.
- 데이터의 수가 매우 적어 train 데이터 기반 모델 예측 성능이 좋다고 판단
- Test 데이터에 대한 성능이 뛰어나지만 본인은 오히려 LGBM의 성능은 현실적이지 않다고 판단하여 XGB 모델을 선택

최고 성능 모델 선정



Confusion Matrix	1	0
1	206	0
0	3	91

	precision	recall	f1-score	support
0	0.99	1.00	0.99	206
1	1.00	0.97	0.98	94
accuracy			0.99	300
macro avg	0.99	0.98	0.99	300
weighted avg	0.99	0.99	0.99	300

Default 파라미터, Grid-Search 파라미터 그리고 하나 임의로 지정한 파라미터 3개 모델을 비교해봤을 때 다음 파라미터를 가진 XGB 모델의 성능이 제일 좋았다.

해당 모델 파라미터 : (n_estimators=300, learning_rate=0.1,max_depth=6)

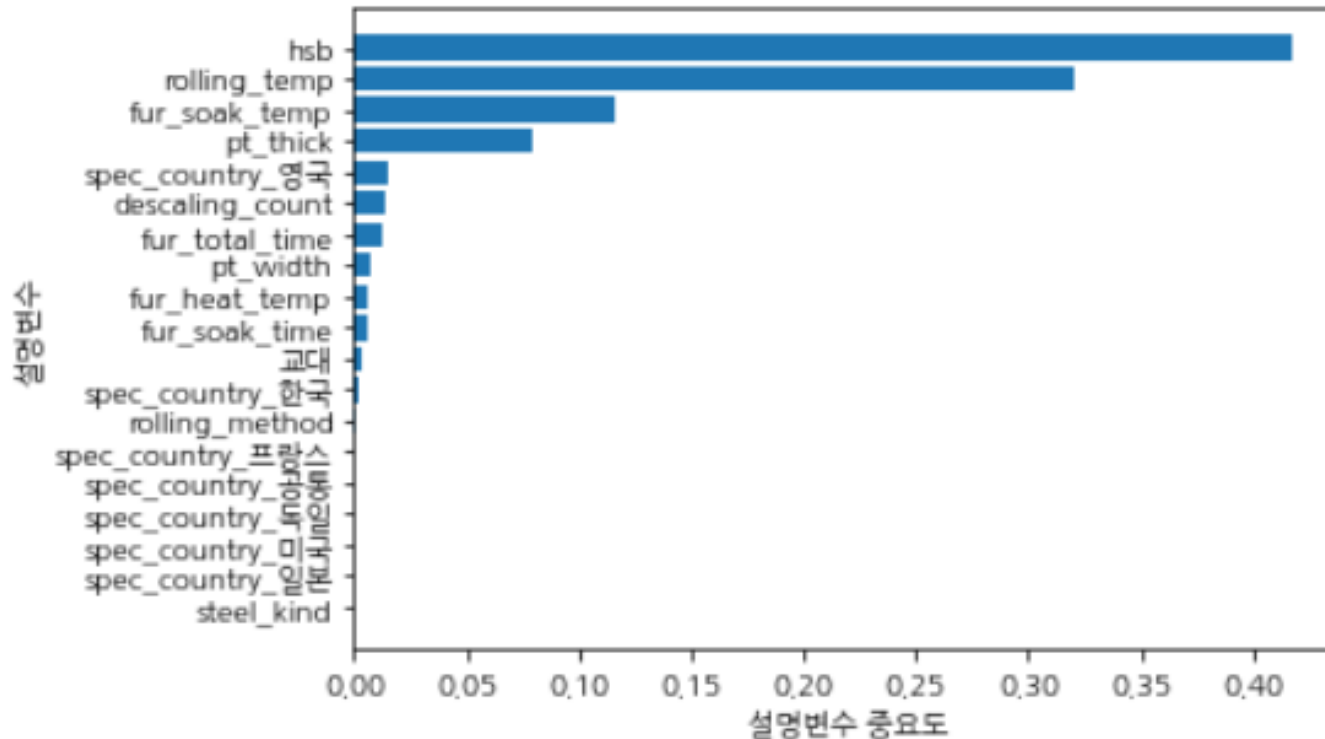
정확도 : 0.99

정밀도 : 1

재현율 : 0.968

F1-Score: 0.984

변수 중요도



- hsb, rolling_temp, fur_soak_temp, descaling_count 등의 중요도가 높게 나왔다. 이 변수들에 대해서는 공정 과정에서 컨트롤 가능한 변수라고 판단했다.
- pt_thick도 비교적 높은 중요도를 가지고 있지만 주문 내역에 기반한 변수이기 때문에 공정 자체에서 컨트롤 가능한 변수가 아니다.
- 결론적으로 온도 조절 및 공정 내에서의 추가 가공을 통해서 불량률을 낮출 수 있다는 판단이 증명됨



05

개선안 도출

개선안 도출



공정 개선안

1. HSB 적용 관련 비용을 산정하고 공정 과정 내에서 적용할 수 있도록 한다.
2. 압연 온도가 1000도를 초과하면 100% 불량 발생하는데 해당 이유를 파악하고 조치하도록 한다.
3. 후판의 두께가 얇을 때 비교적 불량이 발생하기 때문에 해당 규격 공정에 있어서 어떤 애로사항이 있는지 파악한다.
4. 교대 시간대에 따라 유의미한 불량률 차이가 보이기 때문에 근무조별 근무 방식 혹은 근무 공정을 확인한다.

분석 개선 기회 파악

- 도메인 지식이 부족해서 좀 더 깊게 해석하지 못하고 새로운 인사이트를 도출하지 못한 점이 아쉽다.
- 만약 데이터가 더 많았다면 규격별 함량 데이터를 가져와 통합하여 분석해보고 싶다.

The background of the slide is a composite image of an industrial setting. It features a large, blue, curved mechanical component, possibly a part of a conveyor system or a large machine, which is the central focus. The scene is illuminated with a warm, orange glow, likely from industrial lights, creating a hazy and atmospheric effect. In the lower-left corner, there are metal structures and stairs, suggesting a multi-level industrial facility. The overall tone is professional and industrial.

THANK YOU

C1
권구택