

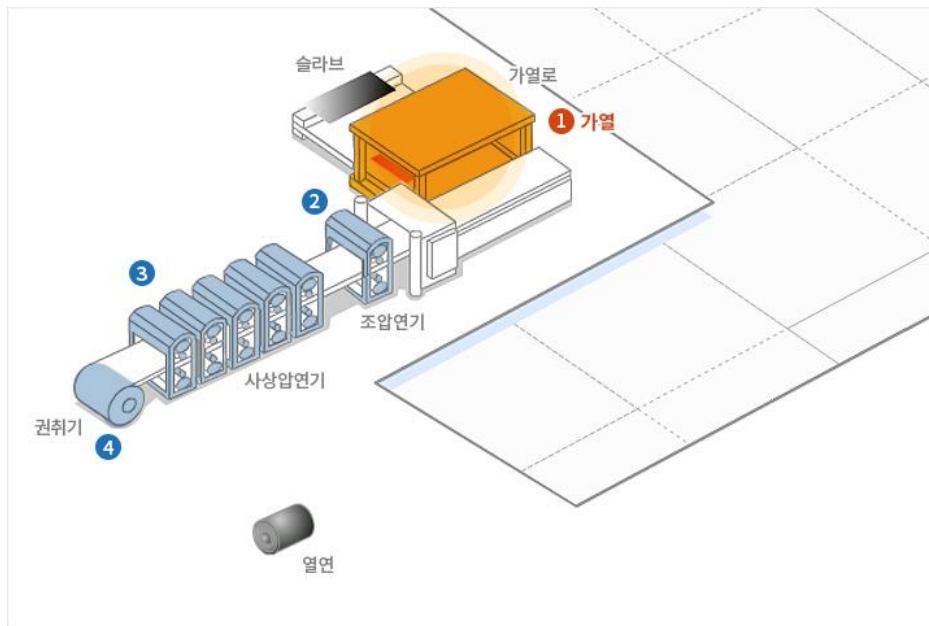
후판 공정 Scale 불량

영향 인자 분석 및 개선안 도출

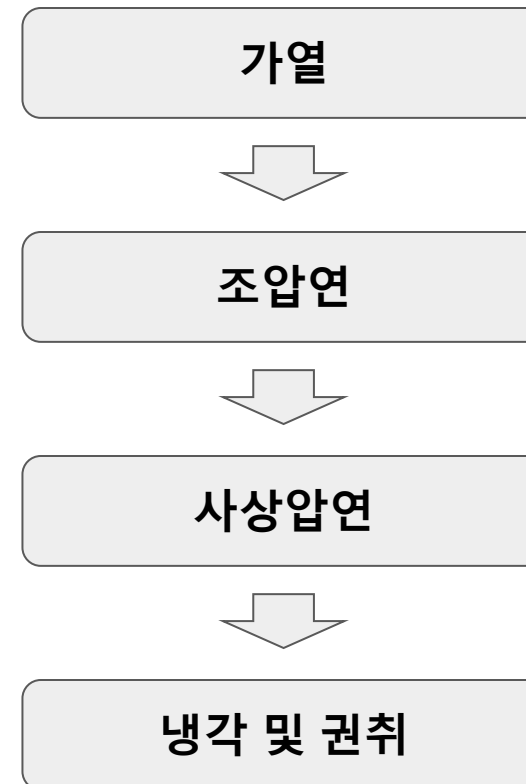
C3조
박경남 김영국 이충현
강민주 이정희 이찬영

추진 배경

최근 들어 선박제조에 주로 사용되는 후판 제품의 “Scale 불량 급증”이라는 이슈가 발생했다. 그 원인을 분석해 본 결과, 압연흄, Scratch등 다양한 불량이 발생했으나 특히 압연공정에서 Scale 불량이 급증한 것을 확인할 수 있었다. 그래서 수집된 데이터를 활용하여 다양한 분석을 통해 불량률의 근본 원인을 찾고 불량 예측 및 개선기회를 도출하고자한다.



<http://product.posco.com/homepage/product/kor/jsp/process/s91p2000120h.jsp>



변수 이름	목표 / 설명변수	데이터 설명
scale	목표변수	Plate번호
plate_no	설명변수	열연작업시각
rolling_date	설명변수	Scale(산화철) 불량
spec_long	설명변수	제품 규격
spec_count ry	설명변수	제품 규격 기준국
steel_kind	설명변수	강종
pt_thick	설명변수	Plate(후판) 지시두께(mm)
pt_width	설명변수	Plate(후판) 지시폭(mm)

변수 이름	목표 / 설명변수	데이터 설명
pt_width	설명변수	Plate(후판) 지시폭(mm)
hsb	설명변수	HSB(Hot Scale Braker)적용여부
fur_no	설명변수	가열로 호기
fur_input_r ow	설명변수	가열로 장입열
fur_heat_te mp	설명변수	가열로 가열대 소재온도(°C)
fur_heat_ti me	설명변수	가열로 가열대 재로시간(분)
fur_soak_t emp	설명변수	가열로 균열대 소재온도(°C)
fur_soak_ti me	설명변수	가열로 재로시간(분)
fur_total_ti me	설명변수	가열로 총 재로시간(분)

변수 이름	목표 / 설명변수	데이터 설명
rolling_method	설명변수	압연방법
rolling_temp	설명변수	압연온도(°C)
descaling_count	설명변수	압연Descaling 횟수
work_group	설명변수	작업조
fur_heat_time	설명변수	가열로 가열대 재로시간(분)
fur_soak_temp	설명변수	가열로 균열대 소재온도(°C)
fur_soak_time	설명변수	가열로 재로시간(분)
fur_total_time	설명변수	가열로 총 재로시간(분)
rolling_method	설명변수	압연방법

변수	파생 변수
spec_long	선체구조용
	교량구조용
	용접구조용
	일반구조용
	보일러압력용기 용
	해양구조용
steel_kind	C(cabon)
	T(titanium)

변수	파생
spec_country	공통
	미국
	영국
	프랑스
	독일
	한국
	일본

변수	파생
hsb	HSB 적용
	HSB 미적용

변수	파생
fur_no	가열로 1호기
	가열로 2호기
	가열로 3호기

변수	파생
fur_input_row	가열로 장입 1열
	가열로 장입 2열

변수	파생
work_group	작업주간(1,3조)
	작업야간(2,4조)

파생 변수	변수설명
fur_combine	1호기 1열
	1호기 2열
	2호기 1열
	2호기 1열
	3호기 1열
	3호기 1열

파생 변수	변수 설명
waiting_time	(fur_total_time - fur_heat_time - fur_soak_time)

01. 데이터 탐색

- 데이터 전처리
 - 이상치 및 결측치
 - 파생변수 생성
- EDA
 - 히스토그램
 - 상관계수

02. 모델 구축

- 로지스틱 회귀분석
- 의사결정나무
- 랜덤 포레스트
- 그래디언트 부스팅
- SVM

03. 적용 및 개선

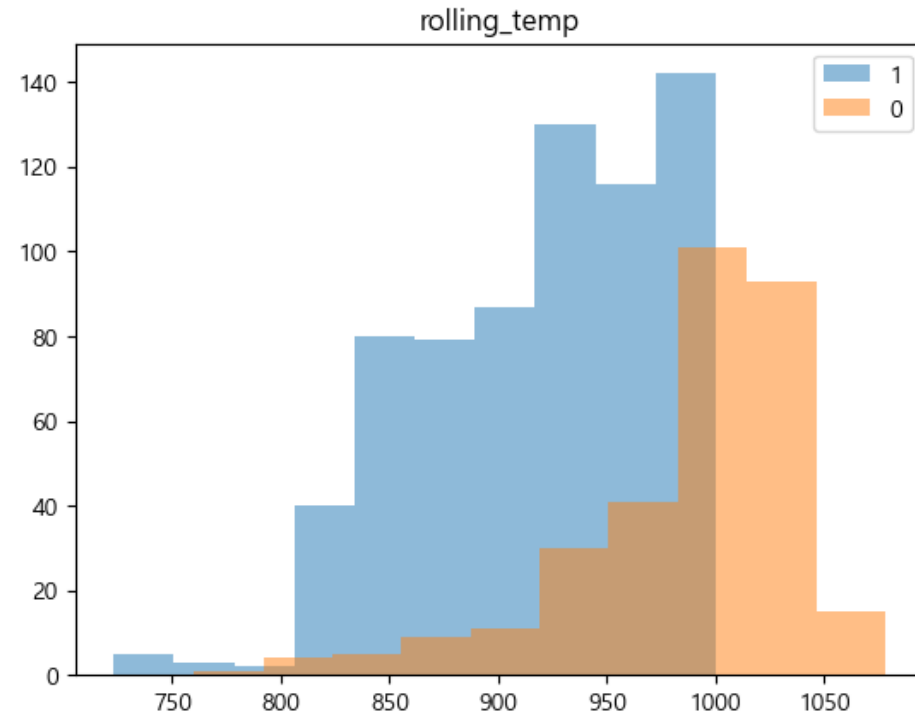
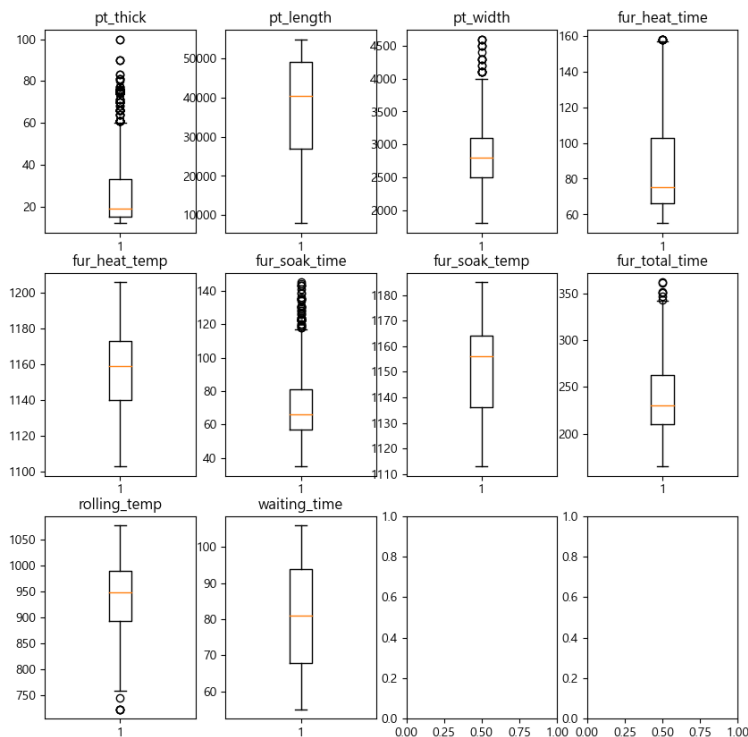
- 알고리즘 모델 평가
- 개선안 도출

목적	분석 계획	
	분석 방법	분석 내용
탐색적 분석	Scatter Plot	두 변수 간의 관계를 시각적으로 표현
	Box Plot	데이터의 분포, 중앙값, 사분위수(IQR), 이상치 등을 시각적으로 요약하는 그래프
	Histogram	연속형 변수의 분포를 막대 형태로 시각화한 그래프
	Heatmap	변수 간 상관관계를 색상으로 표현하는 그래프
분류 모델	다중선형 회귀분석	종속변수와 여러 독립변수 간의 선형적인 관계를 기반으로 예측
	Decision Tree	데이터를 조건에 따라 분할해 예측값을 생성하는 비선형 모델
	Random Forest	여러 개의 결정나무를 앙상블하여 예측 정확도를 향상시키는 모델
	Gradient Boosting	이전 모델의 오류를 보완해가며 순차적으로 모델을 학습시키는 앙상블 기법
	SVM	지지 벡터를 기반으로 클래스 간 마진을 최대화하는 결정 경계를 찾아 분류 성능을 높이는 지도학습 모델

결측치 확인

columns	결측치 개수	결측치 처리	판단 이유
rolling_temp	6	제거	압연 공정 중 0도가 될 수 없다 판단

실제 데이터임을 고려하고, 모든 데이터를 이상치라 보지 않고 상관 관계에 대해 파악하려고 노력했다



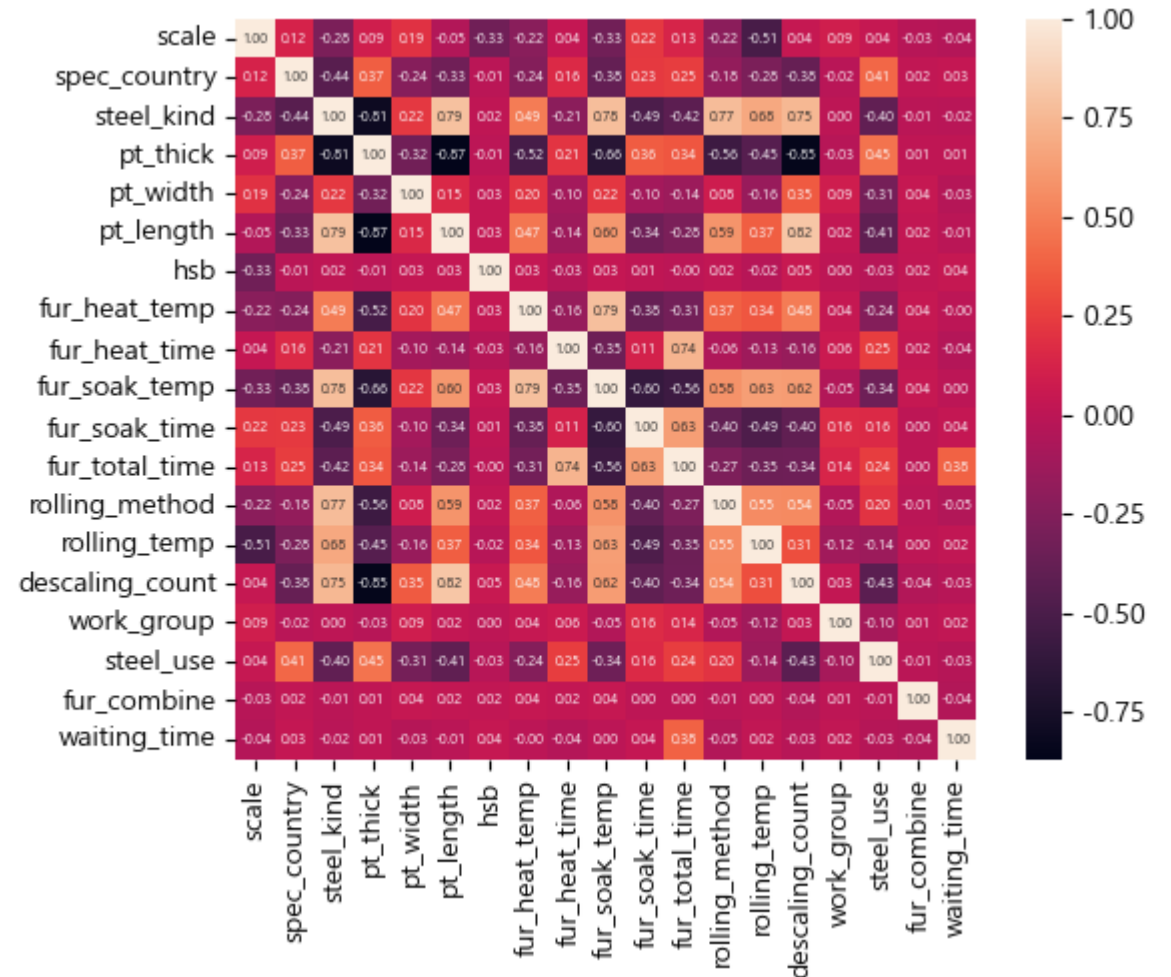
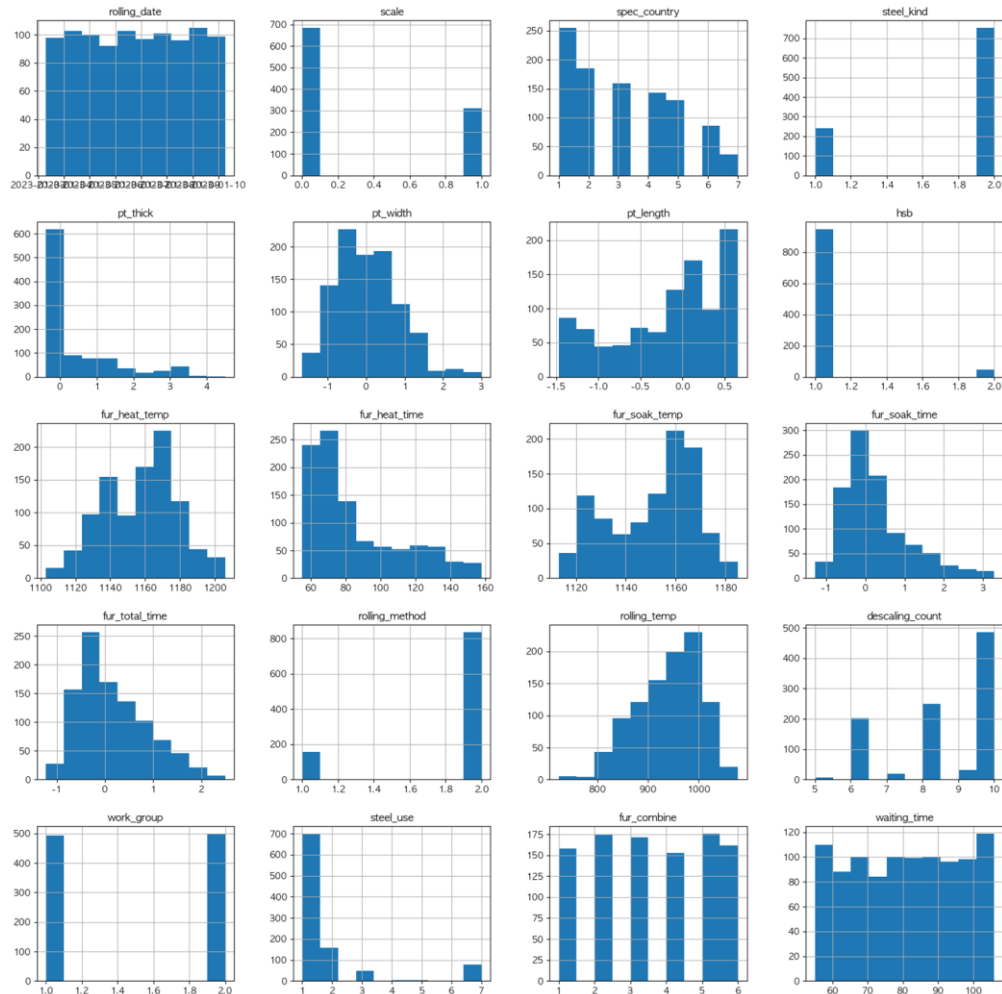
가설 설정

1. descaling count이 불량률에 영향을 미칠 것이다.

2. 압연온도가 불량률에 영향을 미칠 것이다.

3. 가열 후 유지된 온도가 불량률에 영향을 미칠 것이다.

EDA



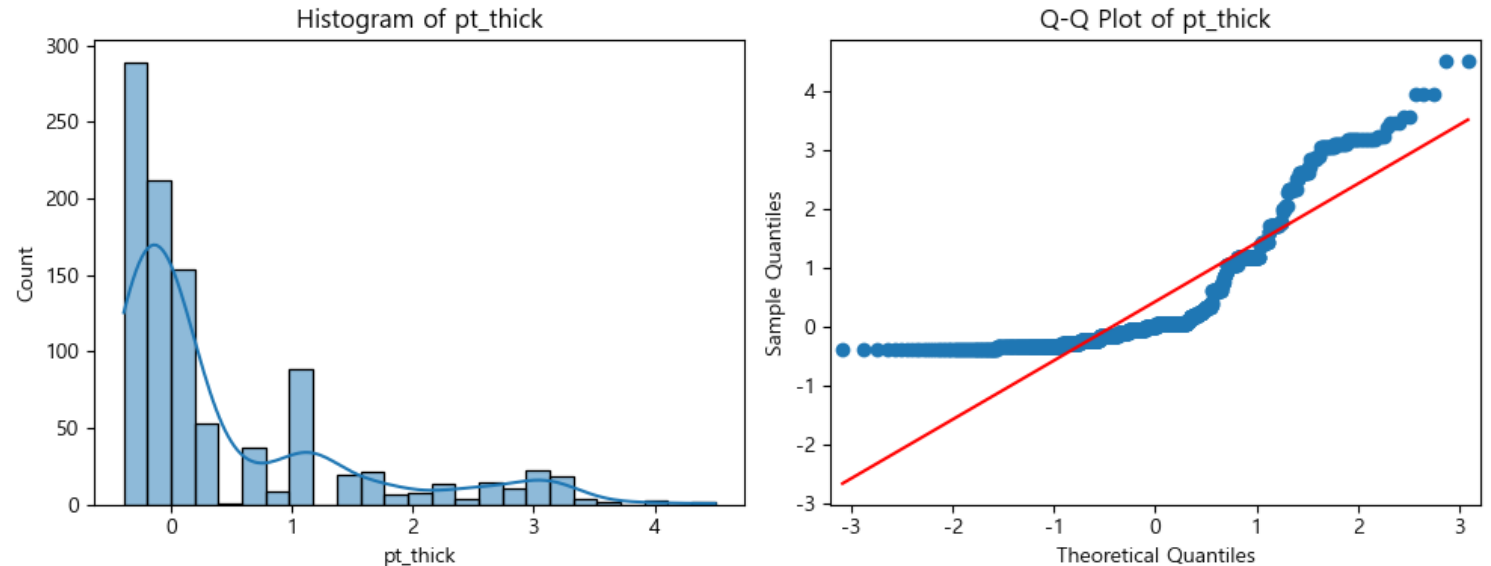
잔차에 대한 가정 검토

등분산성 검토

< 등분산성 검토 (Levene's test) >

pt_thick: statistic = 33.7864, p-value = 0.0000
 pt_width: statistic = 0.2290, p-value = 0.6324
 pt_length: statistic = 88.2880, p-value = 0.0000
 fur_heat_temp: statistic = 16.3428, p-value = 0.0001
 fur_heat_time: statistic = 4.8166, p-value = 0.0284
 fur_soak_temp: statistic = 99.7779, p-value = 0.0000
 fur_soak_time: statistic = 32.2029, p-value = 0.0000
 fur_total_time: statistic = 37.9422, p-value = 0.0000
 rolling_temp: statistic = 17.4189, p-value = 0.0000
 descaling_count: statistic = 9.7080, p-value = 0.0019
 waiting_time: statistic = 0.0358, p-value = 0.8500

정규성 검토



독립성 검토 : 목표변수를 0과 1로 mapping했기에 독립성을 만족한다.

등분산성 검토 : 등분산성을 만족하지 않는다.

정규성 검토 : 데이터가 직선에 적합되어 있지 않기 때문에 정규성을 만족하지 않는다.

카이제곱 검정

< Chi-squared Test Result >

유의한 변수 ($p < 0.05$):

- spec_country: 0.0000
- steel_kind: 0.0000
- hsb: 0.0000
- rolling_method: 0.0000
- work_group: 0.0051
- steel_use: 0.0000

유의하지 않은 변수 ($p \geq 0.05$):

- fur_combine: 0.1751

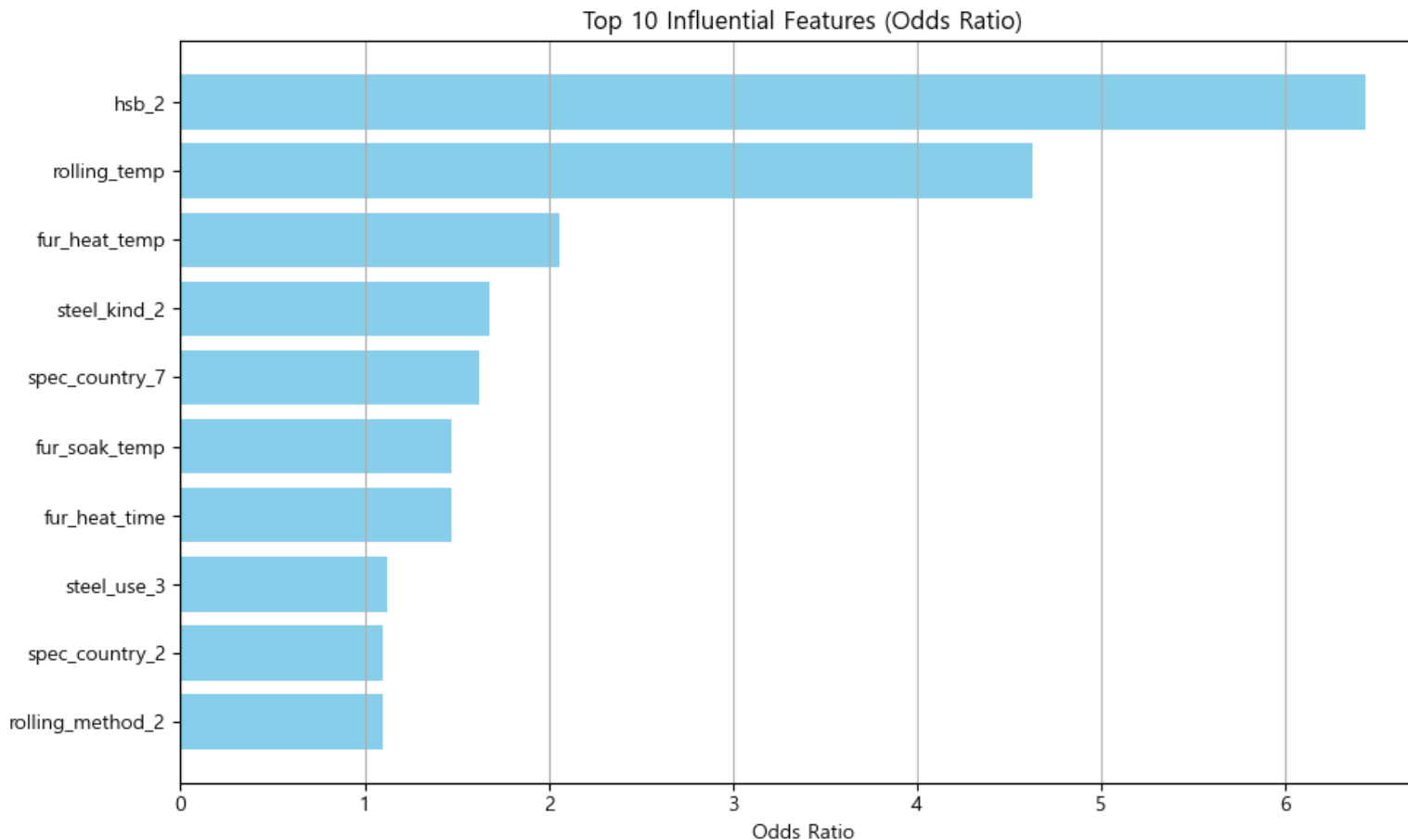
카이제곱 검정 수행 이유

- t-test는 정규성과 등분산성을 전제로 하지만, 해당 가정을 만족하지 않아 신뢰도가 떨어질 수 있으므로 대신 범주형 데이터 간의 관계를 검정하는 카이제곱 검정을 사용하였다.

카이제곱 검정 결과

- spec_country, steel_kind, hsb, rolling_method, work_group, steel_use는 목표변수와 통계적으로 유의미하다.

설명변수의 중요도



최종 결과

학습셋 모델 정확도 0.8817610062893082
테스트셋 모델 정확도 0.8592964824120602

변수 중요도

- **hsb_2, rolling_temp**가 **scale** 발생에
중요한 영향을 미칠 것으로 파악된다

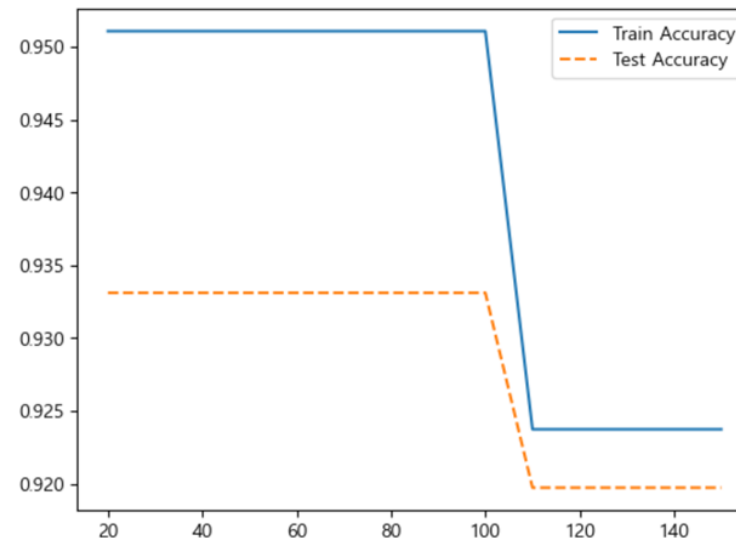
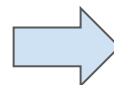
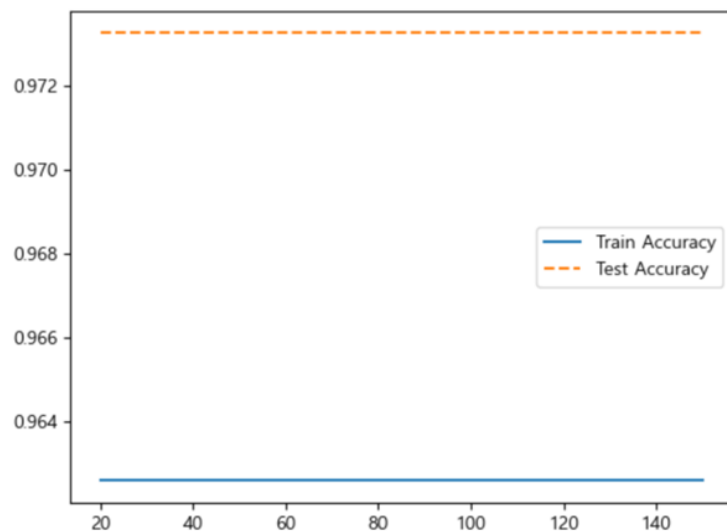
결론

- Train score: 88.2% , Test score:
85.9%로 예측 성능이 안정적인
모델을 확보

데이터분석_의사결정 나무

파라미터 값 선정기준: 1) Train score와 Test score 차이 10% 이내 2) Train score < 98%(과적합 방지)

min_samples_split 파라미터
선정 과정에서 비 이상적인
그래프임을 확인

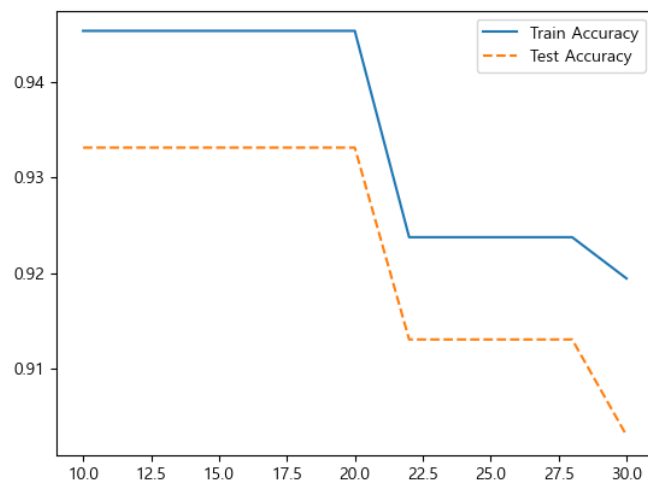


학습 및 평가 데이터 hsb 제거 전

학습 및 평가 데이터 hsb 제거 후

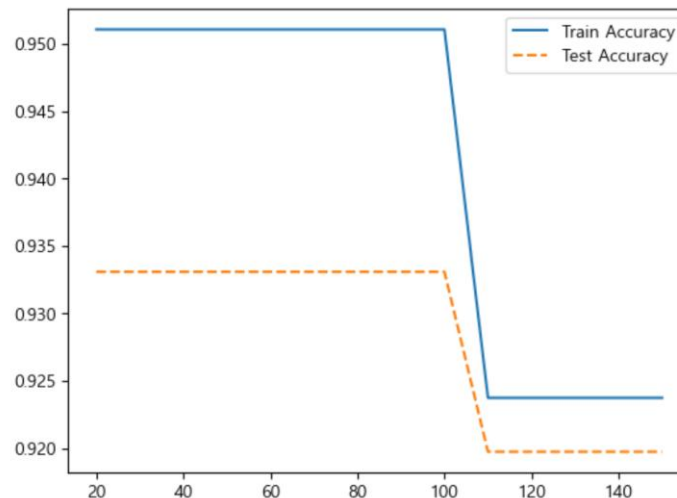
데이터분석_의사결정 나무

파라미터 값 선정기준: 1) Train score와 Test score 차이 10% 이내 2) Train score < 98%(과적합 방지)



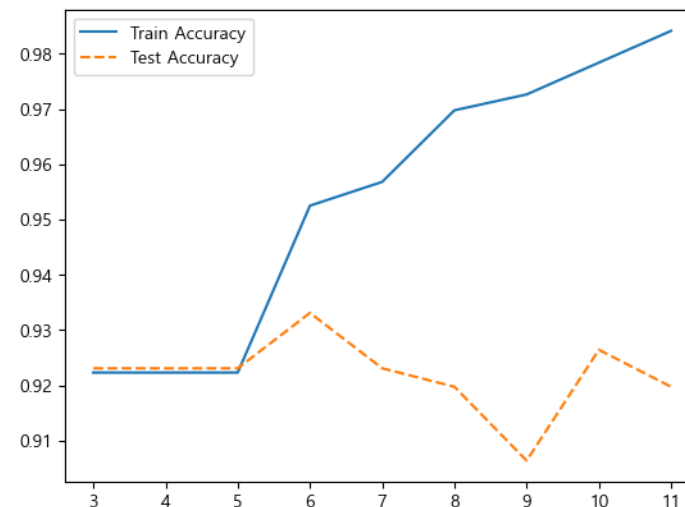
min samples leaf 설정

▶ min samples leaf = 15



min samples split 설정

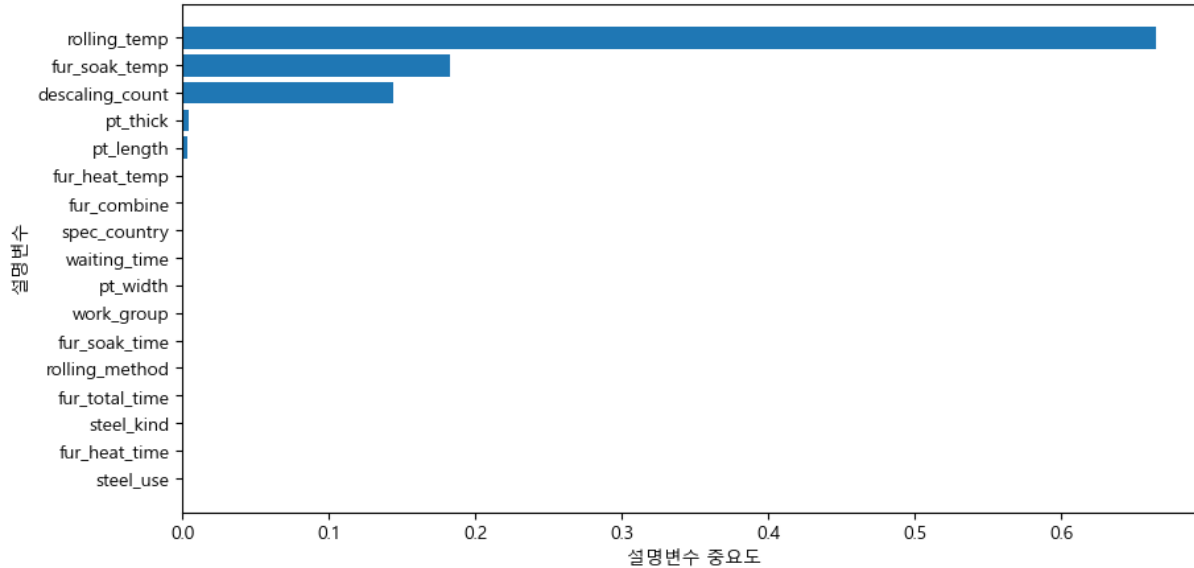
▶ min samples split = 100



max depth 설정

▶ max depth = 6

데이터분석_의사결정 나무



	precision	recall	f1-score	support
0	1.000	0.825	0.904	217
1	0.926	1.000	0.962	478
accuracy			0.945	695
macro avg	0.963	0.912	0.933	695
weighted avg	0.949	0.945	0.944	695

최종 결과

Accuracy on training set: 0.945
Accuracy on test set: 0.933

변수 중요도

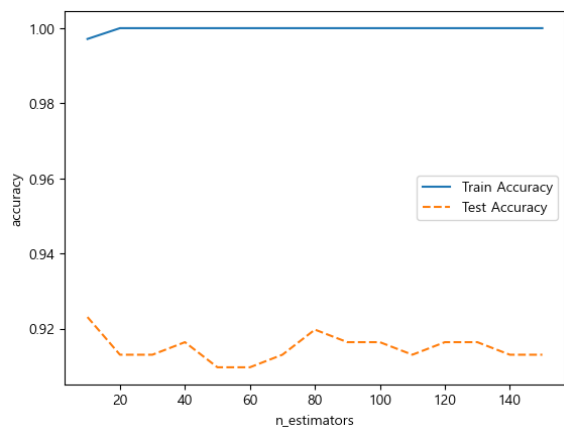
- **rolling_temp**와 **fur_soak_temp**, **descaling_count**는 모델 예측에 가장 큰 영향을 미치는 주요 변수로, 이 변수들을 중심으로 한 예측 모델이 유효한 것으로 확인
- **steel_use** 변수의 중요도는 매우 낮은 것으로 확인

결론

- Train score: 94.5% , Test score: 93.3%로 예측 성능이 안정적인 모델을 확보

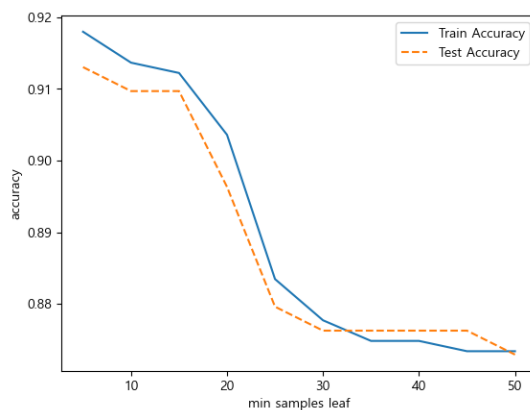
데이터분석_Random forest

파라미터 값 선정기준: 1) Train score와 Test score 차이 10% 이내 2) Train score < 98%(과적합 방지)



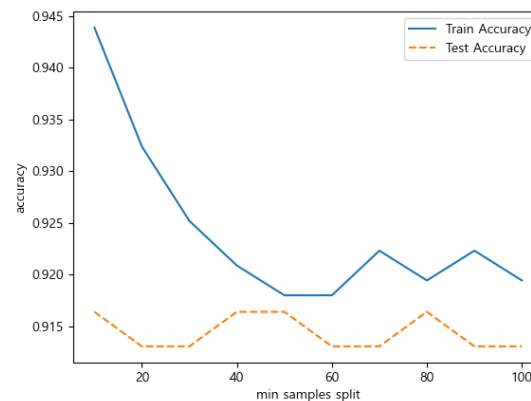
n estimators 설정

▶ n = 40



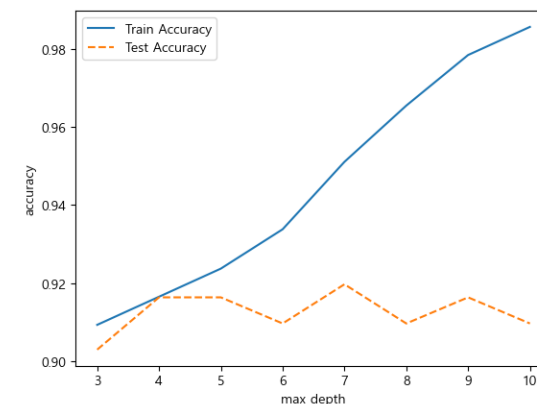
min samples leaf 설정

▶ min samples leaf = 15



min samples split 설정

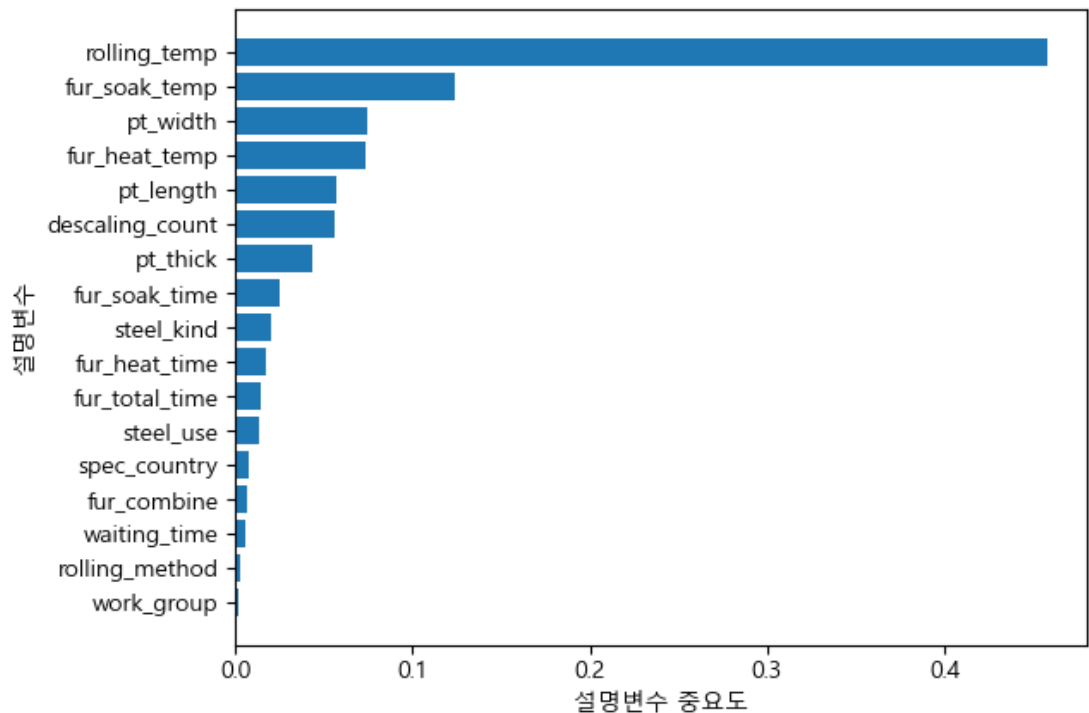
▶ min samples split = 50



max depth 설정

▶ max depth = 7

데이터분석_Random forest



	precision	recall	f1-score	support
0	1.000	0.710	0.830	93
1	0.884	1.000	0.938	206
accuracy			0.910	299
macro avg	0.942	0.855	0.884	299
weighted avg	0.920	0.910	0.905	299

최종 결과

Accuracy on training set: 0.912
Accuracy on test set: 0.910

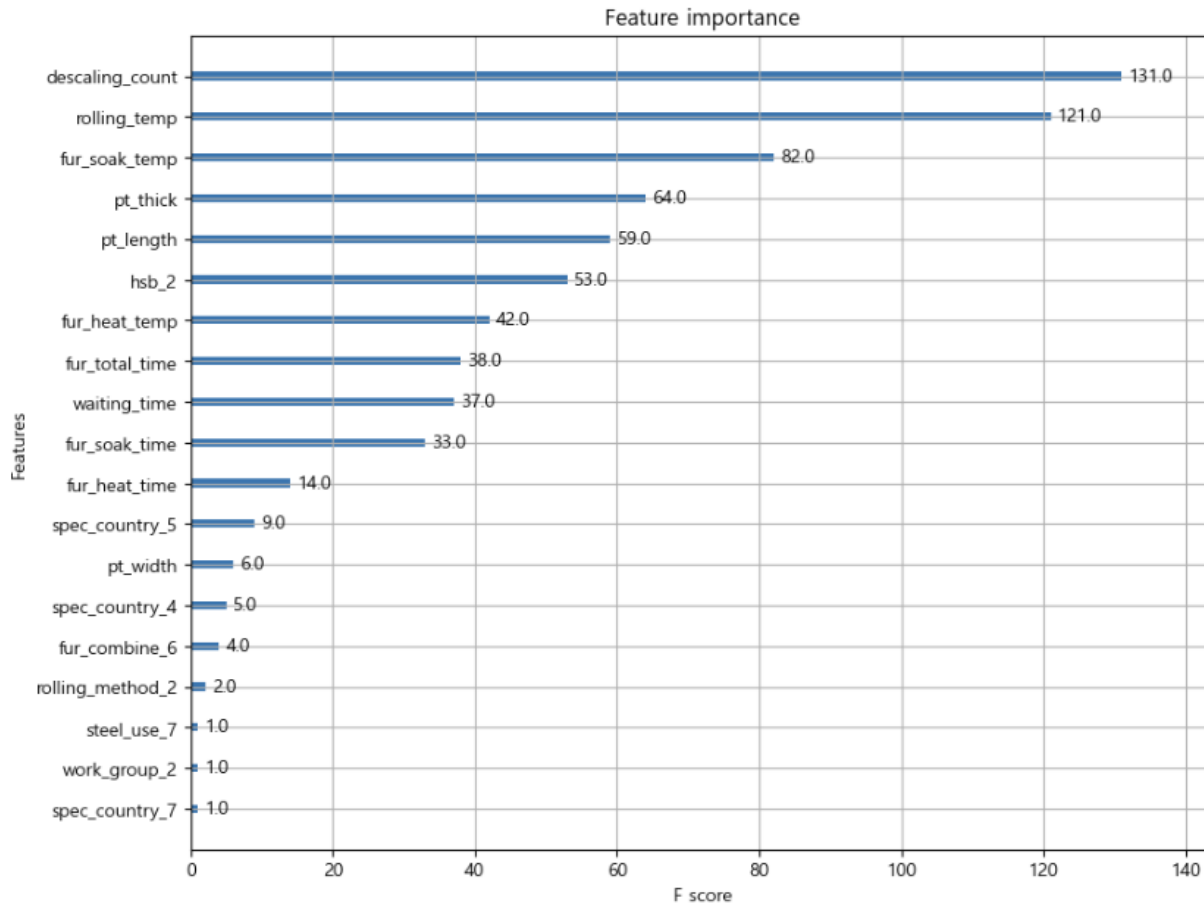
변수 중요도

- **roling_temp**와 **fur_soak_temp**는 모델 예측에 가장 큰 영향을 미치는 주요 변수로, 이 변수들을 중심으로 한 예측 모델이 유효한 것으로 확인
- **work_group** 변수의 중요도는 매우 낮은 것으로 확인

결론

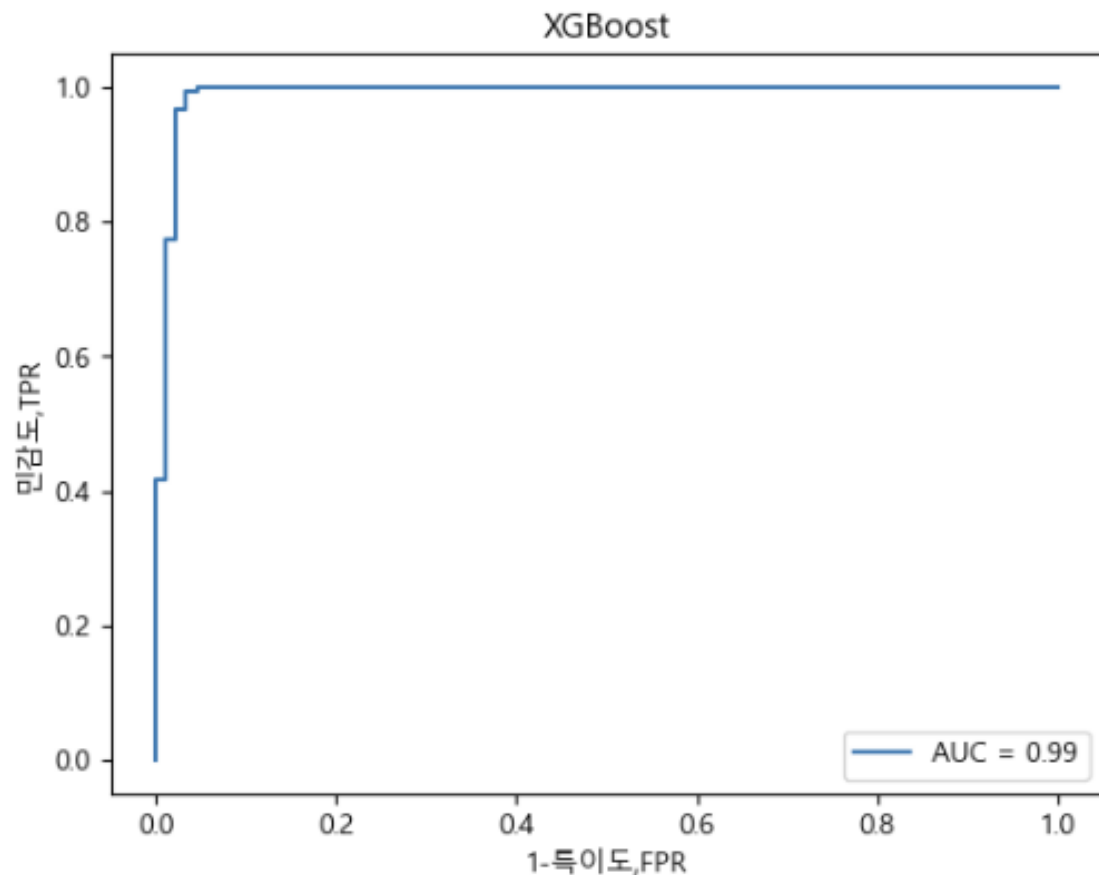
- Train score: 91.2% , Test score: 91.0%로 예측 성능이 안정적인 모델을 확보

데이터분석_XGBoost



- **descaling_count, rolling_temp, fur_soak_temp** 순으로 중요함.
- 주요 변수들이 가열로 및 압연 관련 항목인 것으로 보아 스케일 불량은 공정 온도 및 시간과 밀접한 관계가 있음.

데이터분석_XGBoost



Confusion matrix:

```
[[ 85   3]
 [   1 210]]
```

	precision	recall	f1-score	support
0	0.988	0.966	0.977	88
1	0.986	0.995	0.991	211
accuracy			0.987	299
macro avg	0.987	0.981	0.984	299
weighted avg	0.987	0.987	0.987	299

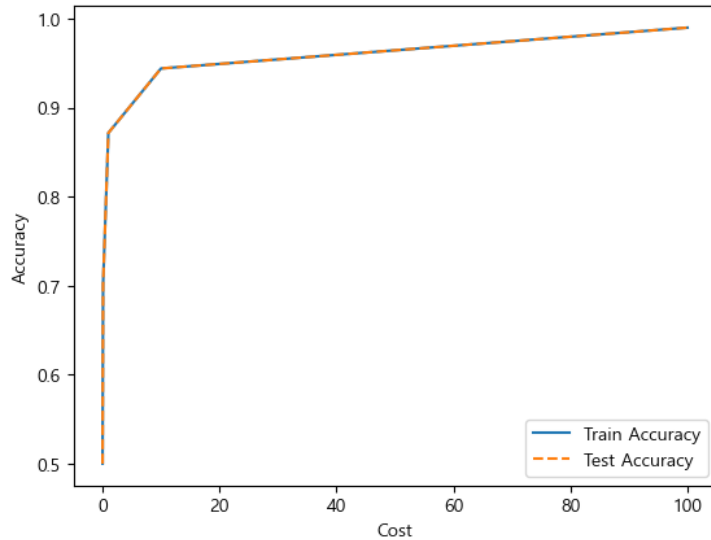
Accuracy on training set: 0.999

Accuracy on test set: 0.987

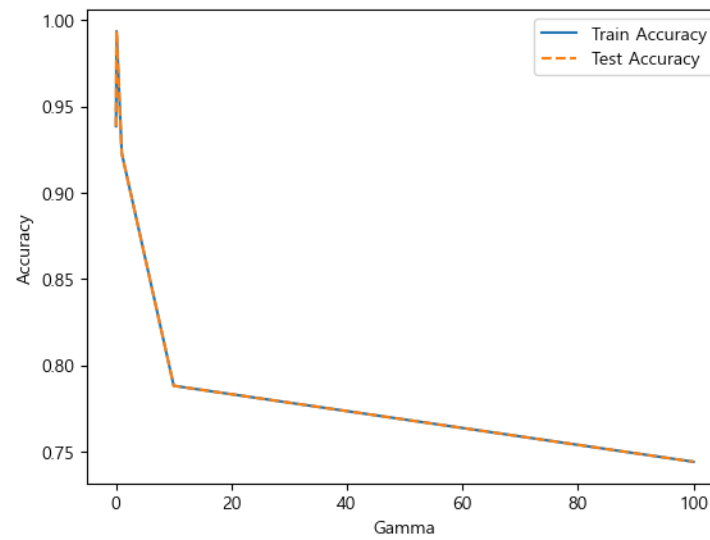
높은 정확도: 학습 데이터 99.9%, 테스트 데이터 98.7%
우수한 분류 성능: Confusion Matrix 및 Precision, Recall, F1-score 결과에서 높은 성능 확인

하이퍼파라미터 선정

그래프 평가



C = 10, gamma = 0.1 Train Accuracy : 0.997
Test Accuracy : 0.997



Grid Search

```
GridSearchCV(cv=10, estimator=SVC(random_state=2021), n_jobs=12,
             param_grid=[{'C': [0.1, 1, 10, 100, 1000], 'kernel': ['linear']},
                          {'C': [0.1, 1, 10, 100, 1000], 'degree': [2, 3, 5],
                           'kernel': ['poly']},
                          {'C': [0.1, 1, 10, 100, 1000],
                           'gamma': ['auto', 'scale'], 'kernel': ['rbf']},
                          {'C': [0.1, 1, 10, 100, 1000],
                           'gamma': ['auto', 'scale'], 'kernel': ['sigmoid']}],
             verbose=1)
```

C = 10, gamma = auto

Train set에 대한 성능 정확도: 0.9780 Test set에 대한 성능 정확도: 0.8593

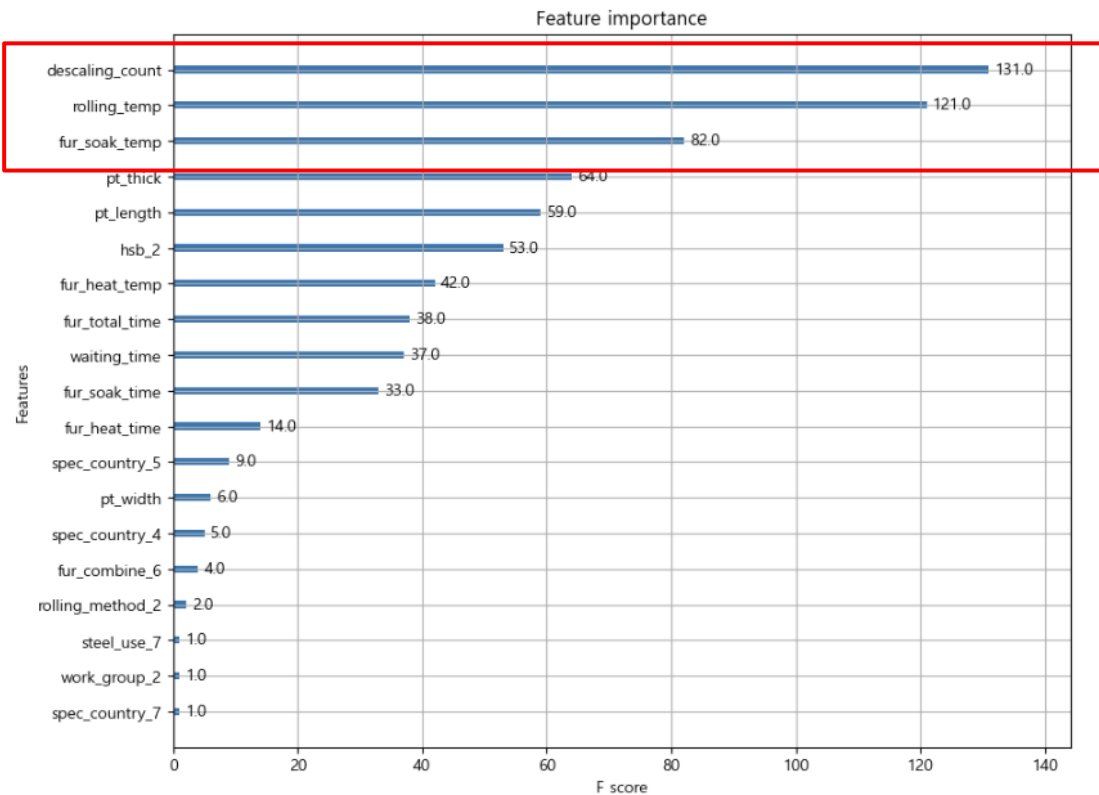
- 그래프 평가와 Grid Search 결과를 서로 비교하면, 그래프 평가의 정확도가 Grid Search의 정확도보다 더 좋으므로 그래프 평가를 사용하는 것이 더 안정적인 모델이라고 할 수 있다.

모델평가

Model	정확도 (Accuracy)		F1-score
	Train	Test	
Logistic Regression	0.882	0.859	0.860
Decision Tree	0.945	0.933	0.944
RandomForest	0.912	0.910	0.905
XGBoost	0.999	0.987	0.987
SVM	0.994	0.994	0.845

- 불균형 데이터가 있는 모델의 성능을 제대로 반영할 수 없기에 F1-score를 사용하여 모델을 평가
- F1-score 결과 가장 성능이 좋은 XGBoost 모델을 사용하는 것이 좋다.

가설 검증



XGboost 변수 중요도

가설 검증 결과 도출

1. descaling count이 불량률에 영향을 미칠 것이다.(O)
2. 압연온도가 불량률에 영향을 미칠 것이다.(O)
3. 가열 후 유지된 온도가 불량률에 영향을 미칠 것이다.(O)

개선안도출

- 1) 실시간 모니터링 기반 제어 최적화
 - Rolling 온도(rolling_temp) 및 균열로 온도(fur_soak_temp)의 편차를 줄이는
 - 실시간 피드백 제어 시스템 구축
- 2) 공정 시간과 디스케일링 횟수 간 균형 제어
 - Total time과 descaling_count의 Trade-off 관계를 고려한 공정 최적화 로직 설계
- 3) 스케일 제거 상태 기반 스마트 디스케일링 제어
 - 스케일 잔존 여부를 실시간 감지하여
 - 디스케일링 횟수·강도를 자동 조정하는 지능형 제어 시스템 도입