

제3회 K-인공지능 제조데이터 분석 경진대회 보고서

프로젝트명	열처리 공정 수율 개선을 위한 공정 파라미터 분석
팀명	Micro Robot
내용요약	<p>1. 문제 정의</p> <ul style="list-style-type: none">- 열처리 제조 공정의 불량 수율 개선을 위한 공정 파라미터 분석- 제품 '불량률'의 제 3사분위수 이상과 이하를 각각 '위험'과 '안정'으로 분류하고, 이를 예측하는 이진 분류 모델을 설계함- 이진 분류 모델을 학습하는데 영향을 준 공정 파라미터를 분석하여, 수율에 영향을 주는 공정 파라미터로 선정 및 관리 <p>2. 제조데이터 정의 및 처리과정</p> <ul style="list-style-type: none">- '공정 데이터'가 21열 * 2,939,722행으로 구성되어 있으며, 레이블 데이터인 '품질 데이터'는 7열 * 136행으로 구성되어 있음- 공정 데이터의 기술 통계량을 데이터로 활용하여 136개인 '품질 데이터' 레이블과 Mapping 함- 학습 데이터가 부족하며, '안정'과 '위험' 레이블의 균형이 맞지 않아 Synthetic Minority Over-sampling Technique을 사용하였고, Standardization으로 데이터 전처리를 진행함 (Train set 70%, Test set 30%로 분할) <p>3. 분석모델 개발</p> <ul style="list-style-type: none">- 데이터 불균형으로 인한 학습 편향을 방지하고자 Minority class에 가중치를 부여하였고, Stratified 5-Fold로 모델 학습을 진행함- 빠른 학습과 결과 분석을 위해 AutoML을 이용하여 15개 모델을 학습 및 비교 분석하였고, Top-1 model로 공정 파라미터 분석을 수행함 <p>4. 분석 결과 및 시사점</p> <ul style="list-style-type: none">- Accuracy 85.4%인 분류 모델을 통해 공정 상태의 '안정' 또는 '위험'을 예측함- Permutation Importance와 Feature Importance 비교를 통해 중요 관찰 파라미터를 선정하고, 예측 결과에 해당하는 파라미터의 중요도를 관찰함으로써 공정 파라미터와 수율간의 관계를 수치적으로 분석 가능함 <p>5. 중소제조기업에 미치는 파급 효과</p> <ul style="list-style-type: none">- 중소제조기업의 공정 엔지니어가 도메인 업무와 병행할 수 있도록 데이터 전처리와 모델 설계를 쉽고 간단하게 진행하였으며, Model Explainer를 이용하여 불량 수율에 영향을 주는 공정 파라미터 분석을 손쉽게 할 수 있음

상기 본인(팀)은 위의 내용과 같이 제3회 K-인공지능 제조데이터 분석 경진대회 결과 보고서를 제출합니다.

2023 년 11 월 03 일

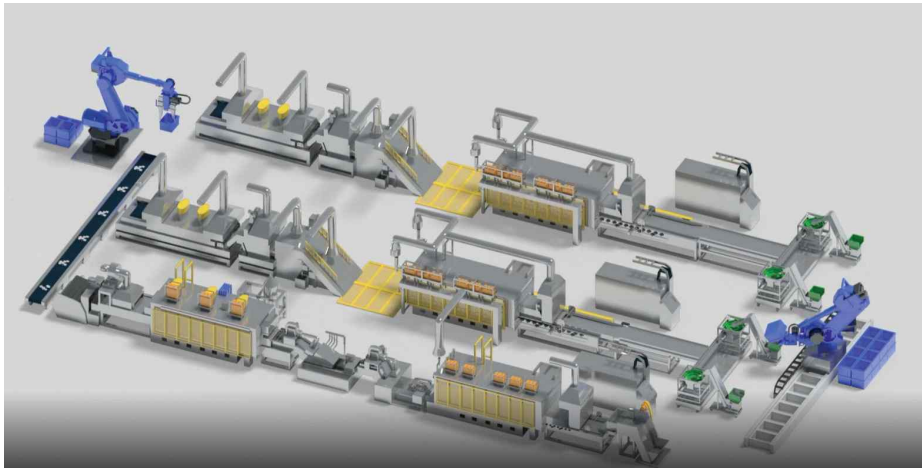
팀장 : 구 자 윤

한국과학기술원장 귀중

□ 문제정의 (HY헤드라인M 15, 줄간격 160)

○ 공정(설비) 개요

- 금속 부품의 열처리(Austempering) 공정으로 제품의 신율, 단면 수축률, 충격치 등의 향상으로 제품 인성을 강화하고 Crack 및 변형 감소를 위해 진행되는 공정이다.



[그림 2] 열처리 공정

○ 이슈사항(Pain Point)

- 2시간 이상 소요되는 무중단 연속공정으로, 공정이 완료되어 불량 유무를 육안으로 보기 전까지 품질 상태를 확인할 수 없다.
- 현재 다양한 공정 데이터가 수집되고 있으나 품질 예측 또는 수율 분석에 활용되지 못하고 현재 설비 상태만 파악 가능하다.

○ 열처리 공정 내 중요 인자 분석을 통한 수율 개선

- 이진 분류 모델을 학습하여 ‘불량단계*’를 예측하고, 모델이 학습한 공정 파라미터의 중요도를 분석하여, 공정 개선점을 도출하는 것을 목표로 한다.

* ‘불량 수량’을 ‘총 수량’으로 나눈 ‘불량률’의 파생 변수

□ 제조데이터 정의 및 처리과정 (HY헤드라인M 15, 줄간격 160)

○ 제조 데이터 특징

- 분야: 자동차 부품 (에어백, 안전벨트)
- 공정명: 열처리 공정 (Austempering)
- 수집기간: 2022년 01월 ~ 2022년 07월 (약 6개월)
- 데이터셋 구조: 테이블 형식 (Tabular Data)
- 데이터 개수:

종류	열 (Column)	행 (Row)	총합 (Total)	파일명 (File)
공정 데이터	21	2,939,722	61,734,162	data.csv
품질 데이터	7	136	952	quality.xlsx

[표 1] 데이터 종류 및 개수

- 주요 변수 정의: '공정 데이터', '품질 데이터', '파생 변수'

종류	설명	데이터형
TAG_MIN	데이터 수집 시간 (1초 간격)	datetime
배정번호	공정 작업 지시 번호-배정번호별 생산	int
건조 1~2존 OP	각 건조 온도 유지를 위한 출력량(%)	float
건조로 온도 1~2 Zone	각 건조로 존의 온도 값	float
세정기	세정기 온도 값	float
소입1~4조 OP	각 소입 존 온도 유지를 위한 출력량(%)	float
소입로 CP 값	침탄 가스의 침탄 능력량(%)	float
소입로 온도 1~4 Zone	각 소입로 존의 온도 값	float
솔트 1~2존 OP	각 솔트존 온도 유지를 위한 출력량(%)	float
솔트 컨베이어 온도 1~2 Zone	각 솔트 컨베이어 존의 온도 값	float
솔트조 온도 1~2 Zone	각 솔트조 존의 온도 값	float

[표 2] '공정 데이터'의 주요 변수 속성 정의 (21개 Column * 2,939,722개 Row 구성)

종류	설명	데이터형
배정번호	공정 작업 지시 번호-배정번호별 생산	int
작업일	공정 일자	datetime
공정명	진행된 공정 이름	object
설비명	작업된 설비 이름	object
양품수량	양품 생산 수량	int
불량수량	불량 발생 수량	int
총수량	전체 수량 (양품 수량 + 불량 수량)	int

[표 3] '품질 데이터'의 주요 변수 속성 정의 (7개 Column * 136개 Row 구성)

종류	설명	데이터형
소요시간	각 배정번호별 작업에 소요된 시간	float
불량률	'불량 수량'을 '총 수량'으로 나눈 파생 변수	float
불량단계 *	'불량률'의 제3사분위수 이상을 '위험', 그 이하는 '안정'으로 분류한 파생 변수이자 종속 변수 (종속 변수)	category

[표 4] 파생 변수'의 주요 변수 속성 정의 (3개 Column * 136개 Row 구성)

* '불량 수량'을 '총 수량'으로 나눈 '불량률'의 파생 변수

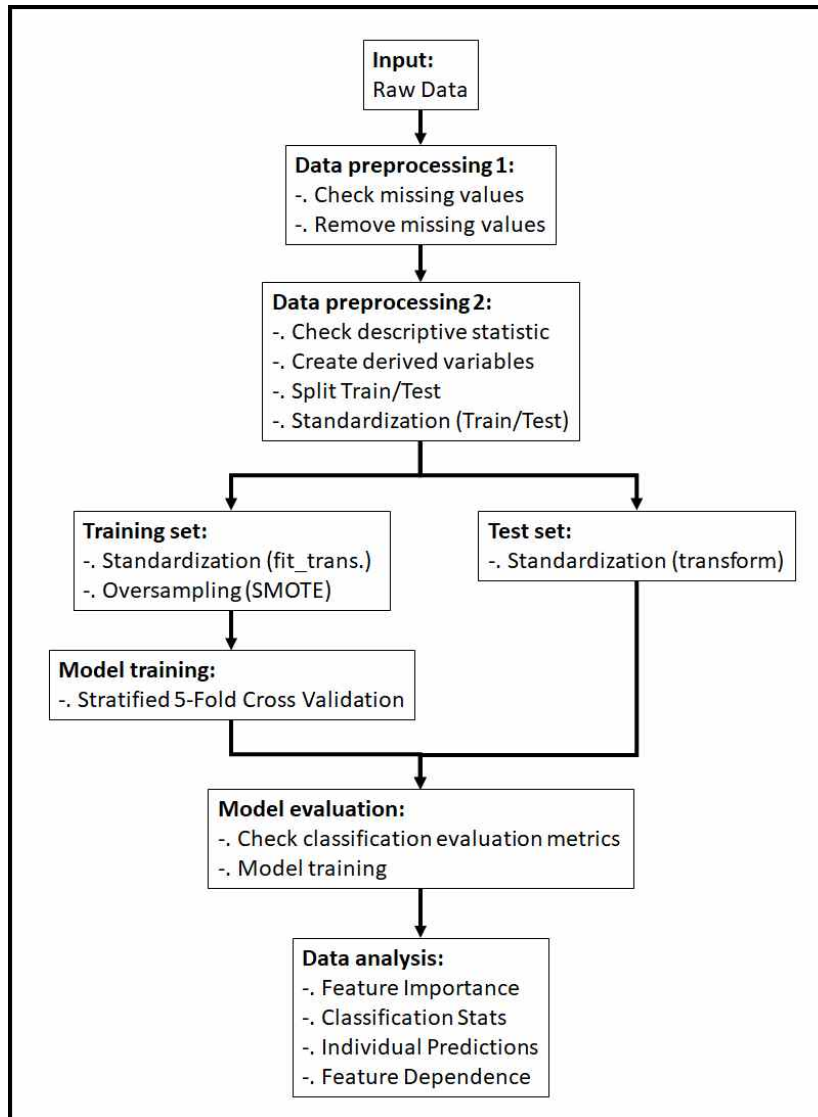
○ 제조 데이터 전처리 과정

- 주요 변수의 기술 통계량 산출: mean, std, skew

변수	개수	평균	표준편차	왜도	결측치
TAG_MIN	2939722	N/A	N/A	N/A	0
건조1존OP	2939721	69.8940	4.0148	-1.1735	1
건조2존OP	2939721	20.4471	5.2171	-1.6585	1
건조로온도1 Zone	2939606	100.0061	0.4360	0.1475	116
건조로온도2 Zone	2939574	100.0198	0.3624	-0.6340	148
세정기	2939631	67.7186	1.6308	-0.2072	91
소입1존OP	2935434	75.6437	25.1608	-1.4246	4288
소입2존OP	2939722	54.8624	4.4291	-1.0886	0
소입3존OP	2939720	53.8603	2.6643	-2.5241	2
소입4존OP	2939719	71.0893	2.5570	-6.7300	3
소입로CP 값	2939721	0.4489	0.0189	-13.5795	1
소입로CP 모니터값	2939575	0.0000	0.0000	0.0000	147
소입로온도1 Zone	2939592	859.2077	3.6477	0.2747	130
소입로온도2 Zone	2939594	860.0021	0.5578	-0.3424	128
소입로온도3 Zone	2939565	860.0029	0.3518	2.3278	157
소입로온도4 Zone	2939552	860.0062	0.4552	19.9666	170
솔트컨베이어온도1 Zone	2939616	283.9963	9.5128	-0.1053	106
솔트컨베이어온도2 Zone	2939580	279.9293	6.6116	-0.0201	142
솔트조온도1 Zone	2939513	331.8062	0.7827	-3.0071	209
솔트조온도2 Zone	2939519	332.1773	0.8733	-2.6661	203

[표 5] ‘공정 데이터’의 주요 변수 기술 통계량 (‘배정번호’를 제외한 20개 Column)

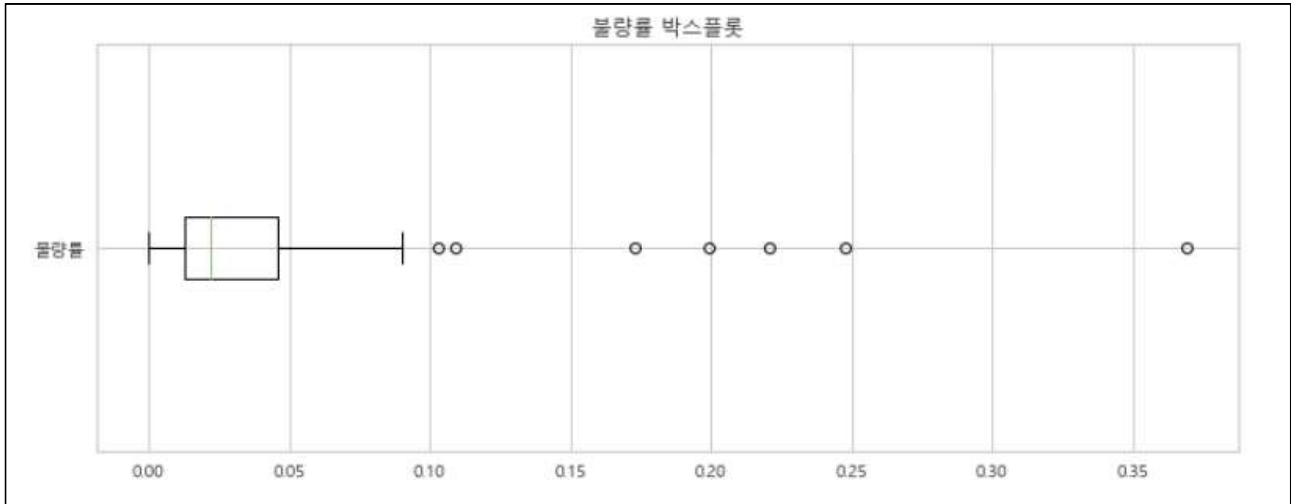
- 배정번호당 생산 물량의 품질 상태가 ‘안정’ 또는 ‘위험’인지 예측해야 하므로 배정번호의 공정 데이터를 기술 통계량으로 요약하였다.
- 결측치는 전체 데이터 수보다 현저히 작으므로 예측 결과에 영향이 없을 것으로 판단되어 제거하였다.
- 기술 통계량으로 ‘평균’, ‘중위수’, ‘표준편차’, ‘왜도’, ‘첨도’, ‘사분위수’, ‘|평균-중위수|’, ‘최소값’, ‘최대값’을 관찰하였으나, 설비 센서의 순간적인 변화(Hunting)가 될 수 있는 ‘최소값’과 ‘최대값’은 제외하였다.
- 경향성을 보여줄 수 있는 통계량 중 ‘평균’, ‘표준편차’, ‘왜도’ 3가지만 데이터로써 사용되었다.



[그림 3] 데이터 전처리 과정 (Flow Chart)

- 입력 데이터로 '공정 데이터'를 사용하였으며, [표 2]에서 'TAG_MIN'과 '배정번호'를 제외한 19개 열의 2,939,722개 행 중 결측치 4,677개 행을 제거하였다.
- 19개 열(공정 데이터)의 기술 통계량*을 추출하였고, '배정번호' 기준 'TAG_MIN'의 시작 시각과 마지막 시각의 차이를 변수로 하는 '소요 시간'을 생성하였다. 하여, 총 58개 열이 데이터 세트로 가공되었다.
* 경향성을 관찰할 수 있는 '평균', '표준편차', '왜도' 만 사용하였다.
- '품질 데이터'에서 [표 4]의 파생 변수인 '불량률'과 '불량단계'를 생성하였고, 가공된 '공정 데이터'의 기술 통계량에 '불량단계'만 결합하여 총 59개 열을 갖는 데이터 세트로 가공하였다.

- ‘불량단계’는 ‘불량률’의 제3사분위 이상은 ‘위험’, 그 이하는 ‘안정’으로 분류한 파생 변수이며, ‘불량률’은 분류기 학습에 영향을 주므로 제거하였다.



[그림 4] “불량률”의 제 3사분위수 이상은 “위험”, 그 이하는 “안정”으로 분류하는 “불량단계” 변수 생성

- 파생 변수인 ‘소요시간’은 배정번호별 소요된 열처리 공정 시간이 서로 다르므로, 불량품 발생에 충분히 관여될 수 있는 파라미터로 판단되어 생성하게 되었다. (공정 시작 시각부터 현재 시각의 차값)

구분	열 (Column)	행 (Row)	생성된 파생 변수	사용 변수
공정 데이터	58	136	‘소요시간’	19개열 * 3(기술통계량) + ‘소요시간’
품질 데이터	2	136	‘불량률’ ‘불량단계’	‘총 수량’ ‘불량단계’
계	60	136	-	-

[표 6] ‘공정 데이터’와 ‘품질 데이터’ 전처리 결과

- 위 전처리 과정을 거쳐, 60개 열과 136개 행을 갖는 최종 데이터 세트로 가공하였고, 분류 모델 학습 및 모델 평가를 위해 학습 데이터와 테스트 데이터로 7:3 비율로 분리하였다.
- 136개 레이블 중 ‘안정’은 105개로 Majority Class이며, ‘위험’은 31개로 Minority Class가 된다. 이러한 데이터 불균형은 모델 학습에 영향을 주므로, Class간 균형이 맞아야 한다.

- 먼저 Training set와 Test set에 대해 Standardization을 적용하였고, Class간 균형을 맞추기 위해 Over-sampling 중 Synthetic Minority Over-sampling Technique을 사용하였다. 이는 동일 모델 Accuracy 기준 Test set에 대해 10% 이상의 성능 향상을 보여주었다.

구분	열 (Column)	행 (Row)	Standardization 적용	SMOTE 적용 후
Training set	59	95	적용(fit_transform)	138
Test set	59	41	적용(transform)	N/A
계	59	136	N/A	N/A

[표 7] Training set와 Test set(Unseen/Holdout) 구성

○ 데이터 품질지수 산출 결과

구분	품질지수	가중치	가중치 지수	오류율
완전성	100.00%	25.00%	25.00%	0.00%
유일성	100.00%	15.00%	15.00%	0.00%
유효성	100.00%	15.00%	15.00%	0.00%
일관성	100.00%	15.00%	15.00%	0.00%
정확성	100.00%	15.00%	15.00%	0.00%
무결성	100.00%	15.00%	15.00%	0.00%
품질 지수	100.00%	100.00%	100.00%	0.00%

[표 8] 데이터 품질지수 산출 결과

□ 분석모델 개발 (HY헤드라인M 15, 줄간격 160)

○ 제조 공정 엔지니어를 위한 쉽고 효과적인 분석 모델 개발

- 실질적으로 중소기업의 제조 공정 엔지니어가 공정 파라미터 분석을 위해 인공지능 모델을 설계하고 현업에 응용하는 일은 쉽지 않다.
- 그럼에도 불구하고, AI 기반의 공정 데이터 분석은 대규모 데이터에 대해 수치적인 예측과 파라미터 설명이 가능하도록 도와주는 강력한 도구이므로 현업에 사용된다면 공정 데이터 분석의 패러다임 변화를 가져올 수 있을 것이다.
- 본 팀은 위와 같은 상황을 고려하여, 제조 공정 엔지니어에게 가장 필요하고 도움이 될 수 있는 AI 기반의 공정 파라미터 분석 방법에 초점을 둔다.
- 공정 엔지니어가 도메인 업무를 수행하면서, 매년 다양한 공정과 상황에 대응하는 AI 모델을 설계하기란 어려움에 가깝다. 때문에, 공정 데이터를 쉽게 가공하고 빠르게 분석하는 분석 모델 설계에 중점을 둔다.

○ AutoML 기반의 이진 분류 모델 설계 및 학습 진행

- 분석 모델: Extra Trees Classifier
- 사용 이유: AutoML을 이용하여 15개 Algorithms을 선정하였고, For 문을 이용하여 하이퍼 파라미터들을 Fine Tuning 하였을 때와 Tuning 하지 않았을 때의 성적을 비교하였다. 이때, 학습 소요 시간이 짧으면서, 상대적으로 타 모델 대비 과적합에 강하고 가장 높은 성적을 보인 모델 Extra Trees Classifier를 선택하게 되었다.

○ 모델 학습 방법

- 데이터 불균형 문제를 해결하고자, 학습 데이터 세트에 한하여 Synthetic Minority Over-sampling Technique(SMOTE)을 적용하였고, 증식 비율은 0.00에서 1.00사이를 0.02씩 증가하였을 때, 가장 좋은 결과가 나오는 범위를 선별하였다.
- 모델 학습 시 Minority class에 가중치를 부여하였다. 가중치는 0.00에서 1.00사이를 For문으로 0.02씩 증가하였을 때, 가장 좋은 성적을 보인 가중치 범위를 사용하였다.
- Stratified K-Fold 방법으로 모델 학습을 하였으며, 데이터 불균형 문제 해소와 일반화 성능 향상을 이유로 사용되었다.
- 상기 과정을 기반으로 약 15개 Algorithms을 For문을 이용하여 학습하였고, Fine Tuning 하였을 때 좋은 성적을 보이는 Top-3 모델 3개를 선별하였다.
- 선별된 모델은 [그림 4]와 같이 RandomForest, XGBoost, Extra Trees였으며, 그 중 Extra Trees는 빠른 연산과 분류 성능 또한 타 모델과 동일한 수준이었기 때문에 분석 모델로 선택 되었다.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Extra Trees Classifier	0.8537	0.7014	0.5556	0.7143	0.6250	0.5358	0.5424

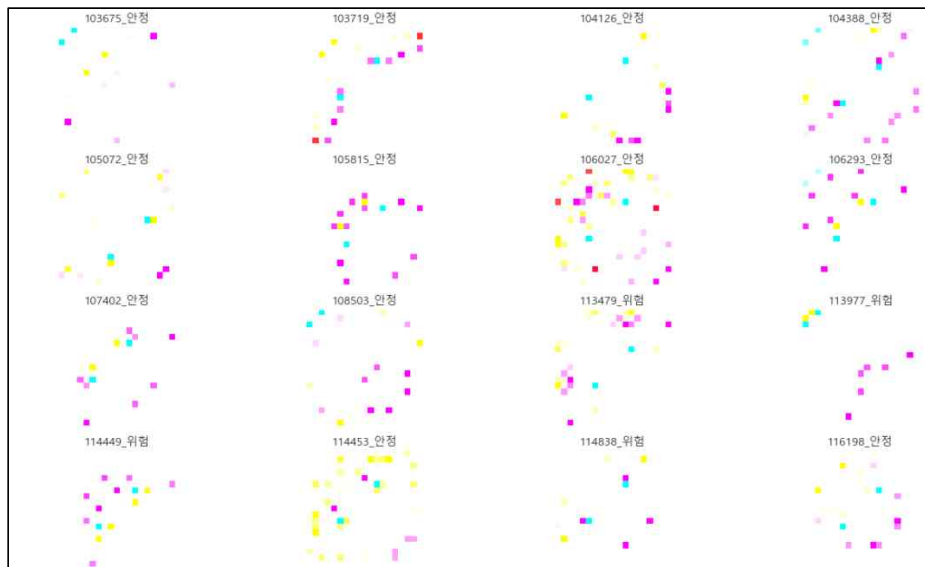
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Extreme Gradient Boosting	0.8537	0.7257	0.5556	0.7143	0.6250	0.5358	0.5424

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Random Forest Classifier	0.8537	0.6632	0.4444	0.8000	0.5714	0.4917	0.5227

[그림 5] Top-3 모델들의 Test set 분류 결과

○ Ensemble 및 CNN 방법을 이용한 분류 모델 설계 결과

- 상기 모델 학습 과정에서 선별된 Top-3 모델을 Ensemble 하였으나, 분류 성능이 단일 모델 대비 좋지 않아 채택되지 않았다.
- CNN을 이용한 이진 분류를 실행해 보았으나 전처리 과정이 복잡해지는 비용 대비, 분류 성능이 월등히 개선되지 않으므로 채택되지 않았다.
- CNN 학습을 위해, 기술 통계량의 상관계수를 이미지로 전처리하여 학습 데이터로 활용하였다. BGR 중, B 값에는 '공정 데이터'의 기술 통계량 값이 들어갔고, G 값에는 '공정 데이터'를 절댓값(.abs())으로 변환 후 기술 통계량을 추출하였다. R 값에는 '공정 데이터'의 차이(.diff()) 값의 기술 통계량을 추출하여 입력하였다.



[그림 6] '공정 데이터'의 상관계수를 이미지화한 모습

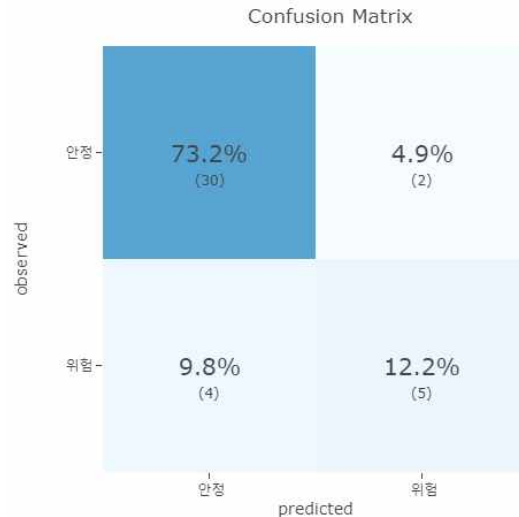
	precision	recall	f1-score	support
안정	0.86	0.83	0.84	23
위험	0.33	0.40	0.36	5
accuracy			0.75	28
macro avg	0.60	0.61	0.60	28
weighted avg	0.77	0.75	0.76	28

[그림 6] CNN Model 기반의 '안정' / '위험' 이진 분류 결과

□ 분석결과 및 시사점 (HY헤드라인M 15, 줄간격 160)

○ 이진 분류 모델의 학습 결과 및 분류 성능 지표

- Training set와 Test set의 비율은 7:3이며, 학습 결과와 분류 성능 지표는 하기 [표 9]과 같다.
- 학습 데이터는 앞에서 언급하였듯이 SMOTE를 사용하여, Minority class인 '위험' 데이터 개수를 Majority class인 '안정' 데이터 수 만큼 생성하였고, 그 학습 결과, SMOTE를 사용하지 않았을 때보다 모델 성능이 향상되었다.
- [표 10]의 학습 결과를 보면 모델이 Test set을 일반화하지 못하는 과적합 문제를 보이지만, 주어진 전체 데이터 수가 136개로 학습 데이터가 부족하였고 Class간 데이터 불균형 문제 또한 존재하였기 때문에 Training set와 Test set간에 차이가 발생한 것으로 보인다. 이와 같은 문제는 공정 데이터가 충분히 모인다면 해결 가능한 문제로 보인다.



[그림 8] 최종 분석 모델인 Extra Trees Classifier의 예측 결과 (Accuracy 85.4%)

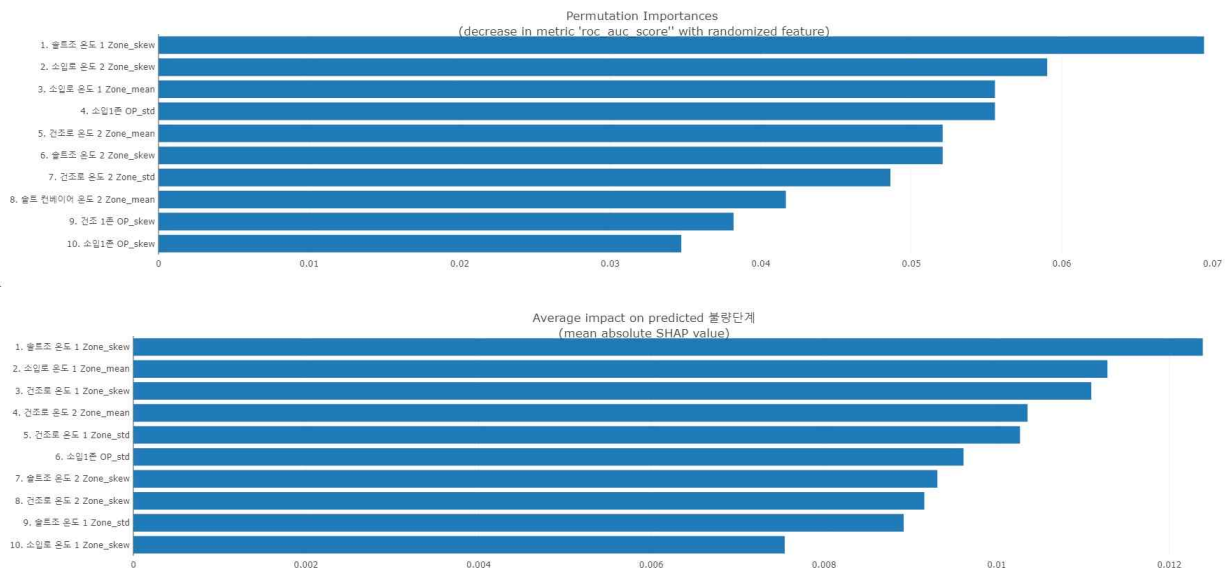
	Precision	Recall	F1-Score	Support
안정	0.88	0.94	0.91	32
위험	0.71	0.56	0.63	9
Accuracy	N/A	N/A	0.85	41
Macro avg	0.80	0.75	0.77	41
Weighted avg	0.85	0.85	0.85	41

[표 9] Extra Trees Classifier의 분류 성능 지표 결과

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Training set	0.8479	0.9590	0.9560	0.8010	0.8667	0.6958	0.7222
Test set	0.8537	0.7014	0.5556	0.7143	0.6250	0.5358	0.5424
Gap	0.0058	0.2576	0.4004	0.0867	0.2417	0.16	0.1798

[표 10] Extra Trees Classifier를 학습한 결과와 Test set 분류 결과 비교

- **Permutation/Feature Importance** 분석 기반의 중요 공정 파라미터 선별
 - ‘불량단계’를 분류하는 모델 학습에 영향을 주었던 공정 파라미터의 중요도를 분석함으로써 중요 공정 파라미터를 선별하고자 한다.
 - 하기 [표 11]은 Permutation Importance와 Feature Importance 결과의 Top-10이며, 양측 모두 ‘솔트조 온도 1 Zone_skew’가 Top-1으로 나왔다.
 - 사용된 기술 통계량은 ‘skew’, ‘mean’, ‘std’이며, 이 중 공정 데이터를 표현함에 있어서 ‘skew’가 미세하게 우위에 있는 것으로 보이지만, 전체 파라미터를 두고 보았을 때 뚜렷한 강세는 보이지 않는다.



[그림 9] Permutation/Feature Importance 결과

NO.	Permutation Importances	Feature Importances
1	솔트조 온도 1 Zone_skew	솔트조 온도 1 Zone_skew
2	소입로 온도 2 Zone_skew	소입로 온도 1 Zone_mean
3	소입로 온도 1 Zone_mean	건조로 온도 1 Zone_skew
4	소입1존 OP_std	건조로 온도 2 Zone_mean
5	건조로 온도 2 Zone_mean	건조로 온도 1 Zone_std
6	솔트조 온도 2 Zone_skew	소입1존 OP_std
7	건조로 온도 2 Zone_std	솔트조 온도 2 Zone_skew
8	솔트 컨베이어 온도 2 Zone_mean	건조로 온도 2 Zone_skew
9	건조 1존 OP_skew	솔트조 온도 1 Zone_std
10	소입1존 OP_skew	소입로 온도 1 Zone_skew

[표 11] Permutation Importance와 Feature Importance의 Top-10 결과

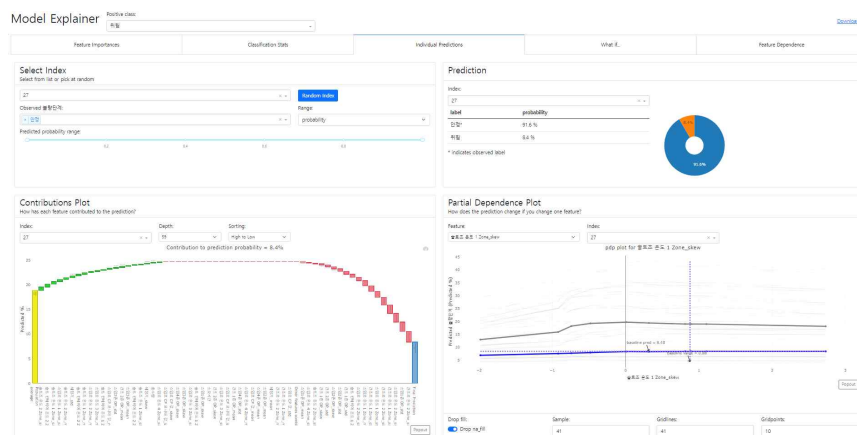
- 기술 통계량 값을 제외하고 공정 파라미터 종류만 본다면, [표 12]와 같이 ‘솔트조 온도 1 Zone’, ‘소입로 온도 1 Zone’, ‘건조로 온도 2 Zone’이 모델 예측에 영향을 준 중요 파라미터 보여진다.

종류	설명	데이터형
솔트조 온도 1 Zone	솔트조 존의 온도 값	float
소입로 온도 1 Zone	소입로 존의 온도 값	float
건조로 온도 2 Zone	건조로 존의 온도 값	float

[표 12] 선별된 공정 파라미터 용도 (도메인 지식이 없는 이상 자세히 알 수 없음)

○ 공정 파라미터 분석 결과 (시사점)

- Permutation Importance는 모델 훈련 후 각 특성의 중요도를 평가하는데 사용되며, 이를 기반으로 [표 11]의 결과를 보았을 때, ‘솔트조 온도 1 Zone’과 ‘소입로 온도 2 Zone’, ‘건조로 온도 2 Zone’ 등의 공정 파라미터가 모델 성능에 영향을 주었으므로 주요 관찰 대상으로 분석된다.
- 열처리 공정 엔지니어는, 상기 결과를 바탕으로 [표 12]에 해당하는 공정에 온도 센서나 제품 특성과 관련된 파라미터를 관찰할 수 있는 센서를 추가 장착하는 것을 고려해볼 수 있다. 또는 ‘열처리_염욕_1’ 설비 외 다른 설비 데이터에도 동일한 방법을 전개하여 설비상의 차이점을 비교 분석해 볼 수 있다.
- 더 자세한 분석은 Model Explorer를 이용하여, 분류 모델 성능에 영향을 준 파라미터의 기여도를 확인해 볼 수 있다. 도메인 지식을 가지고 있는 열처리 공정 엔지니어가 참고한다면, 공정상의 문제를 추정하고 가설을 설립하는데 많은 도움이 될 것이다.



[그림 10] 분류 결과에 영향을 준 공정 파라미터의 기여도(중요도) 분석을 통하여 관찰이 필요한 공정 파라미터 선정이 가능하다.

□ 중소제조기업에 미치는 파급효과(HY헤드라인M 15, 줄간격 160)

○ 공정 파라미터 중요도 분석을 통한 설비(공정) 개선점 확인

- 제조 라인에 상주하면서 도메인 업무를 수행하는 실무자들은 다양한 유형의 공정 데이터 분석을 위해 매번 모델을 설계하고 깊은 수준의 데이터 전처리를 수행하는 일은 쉽지 않다.
- 본 팀은 실제 공정 엔지니어들이 기존 도메인 업무와 병행 가능한 수준에서 쉽고 효과적으로 기계학습을 이용하여 공정 개선점을 분석 및 도출하는 것을 목표로 프로젝트를 수행하였다.
- 열처리 공정에 대한 도메인 지식이 없음에도 불구하고, 그리고 학습 가능한 데이터 수가 부족한 상황에서도 공정 상태의 '안정'과 '위험' 예측을 납득 가능한 수준으로 예측하였고, 모델 학습에 영향을 주는 공정 파라미터의 중요도 분석을 통해 설비 또는 공정상에 개선점 도출까지 완료하였다.
- 이와 같이, 간단한 전처리와 기계학습 과정만으로도 공정 데이터 분석이 가능하였으므로, H/W 데이터와 품질 계측 데이터 수집이 가능한 중소제조기업에서는 충분히 효과적으로 사용 가능할 것으로 보인다.