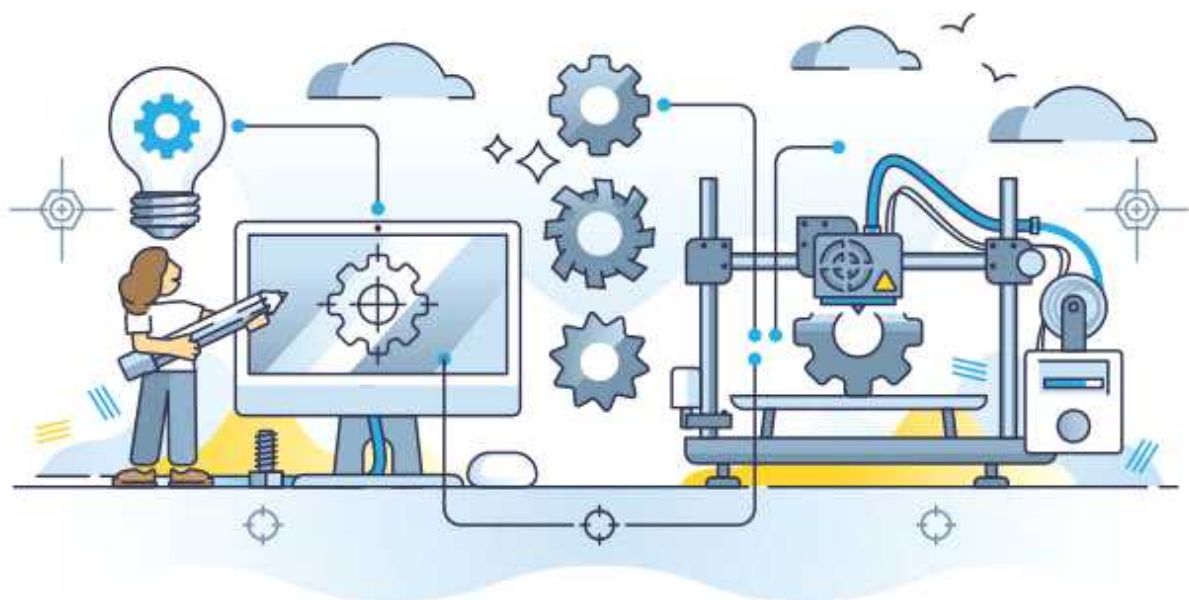


「지역특화 제조데이터 활성화사업」

제조데이터 가이드북

사출공정 데이터셋을
활용한 검사공정
AI 적용 가이드북



Contents

① 분석요약	1
② 분석개요	1. 분석배경	2
	가. 공정(설비) 개요	
	나. 이슈사항	
	2. 분석 목표	6
	가. 분석 목표	
	나. 데이터 정의 및 소개	
	다. 데이터 분석 기대효과 및 시사점	
③ 분석방법	1. 제조데이터 소개	8
	가. 데이터 수집 방법	
	나. 데이터 유형/구조	
	다. 데이터 품질 전처리 및 품질지표	
	2. 분석 모델 소개	14
	가. 데이터 흐름 및 인공지능 모델 적용 흐름도	
	나. AI 분석 모델	
④ 분석결과	1. 제조 데이터셋 수집 및 분석 결과	18
	가. 제조 데이터셋 수집 내역	
	나. 제조 데이터셋을 활용한 분석 결과	
	2. 타현장 적용 방안	21
	가. 유사 타현장의 「AI 데이터셋」 분석 적용 방안	

No	구분		내용
1	분석목적 (현장이슈, 목적)		<ul style="list-style-type: none"> - 현장이슈 : 품질검사는 자동차 부품 생산에서 상당히 중요한 공정임. 작은 불량도 단순 제품 결함으로 끝나는 것이 아니라, 대형 인명피해로 이어질 수 있음. 그러므로 높은 검사정확도를 요구함. 하지만, 작업자의 경험에 의한 육안검사 방식은 부정확한 검사 및 관리지표 체계화 부재 등의 한계점이 있음 - 목적 : 사출공정에서 발생하는 온도, 압력, 속도 등 다양한 데이터를 활용하여 생산품의 품질을 분류하는 AI모델을 만드려함
2	데이터셋 형태 및 수집방법 (csv, json, image 파일 등)		<ul style="list-style-type: none"> - 데이터셋 형태 : 테이블 형식(tabular) 데이터 - 수집방법 : MES에 저장되는 사출기 데이터를 MongoDB에 저장 - 데이터셋 파일 형식(확장자) : .csv - 주요 수집 데이터 : 호퍼온도, 배럴온도, 금형온도, 사출시간 등 사출공정 데이터
3	데이터 개수 데이터셋 총량		<ul style="list-style-type: none"> - 데이터 개수 : 11,280개 - 데이터셋 총량 : 15.981mb
4	분석적용 알고리즘	알고리즘	Decision Tree (의사결정나무) Random Forest (랜덤포레스트) Adaboost (에이다부스트)
		알고리즘 간략소개	Decision Tree는 여러 규칙을 적용하여 독립변수의 공간을 분류 분석하는 알고리즘임. Random Forest는 Decision Tree의 여러 단점들을 극복하여, 훈련과정에서 구성된 다수의 결정 트리로부터 분류분석을 진행하는 알고리즘임. Adaboost는 부스팅 계열 알고리즘으로, 약한 분류기들이 상호보완하며 순차적으로 학습하고, 학습한 것을 조합하여 최종적으로 강한 분류기의 성능을 향상시키는 알고리즘임
5	분석결과 및 시사점		<ul style="list-style-type: none"> - 분석결과 : 본 가이드북은 사출공정에서 수집된 데이터들을 여러 의사결정나무 기반 머신러닝 알고리즘들을 활용하여 품질 분류 모델을 개발함. 분석결과, Random Forest모델이 사출공정의 품질 분류에 가장 적합한 모델임을 확인하였고, 그 정확도 역시 타 모델 대비 가장 높음을 확인하였음 - 시사점 : 사출공정에서 수집된 데이터를 활용해 품질분류 AI 모델을 만들어 적용시, 사출공정의 품질 향상과 비용절감에 기여할 수 있을 것으로 기대됨

1. 분석 배경

가. 공정(설비) 개요

□ 사출 성형의 정의

- 200도 전후의 비교적 높은 온도에서 합성 수지(플라스틱)등의 재료를 녹여 금형에 주입한 뒤 냉각시켜 원하는 제품으로 성형해 만드는 제조 방법

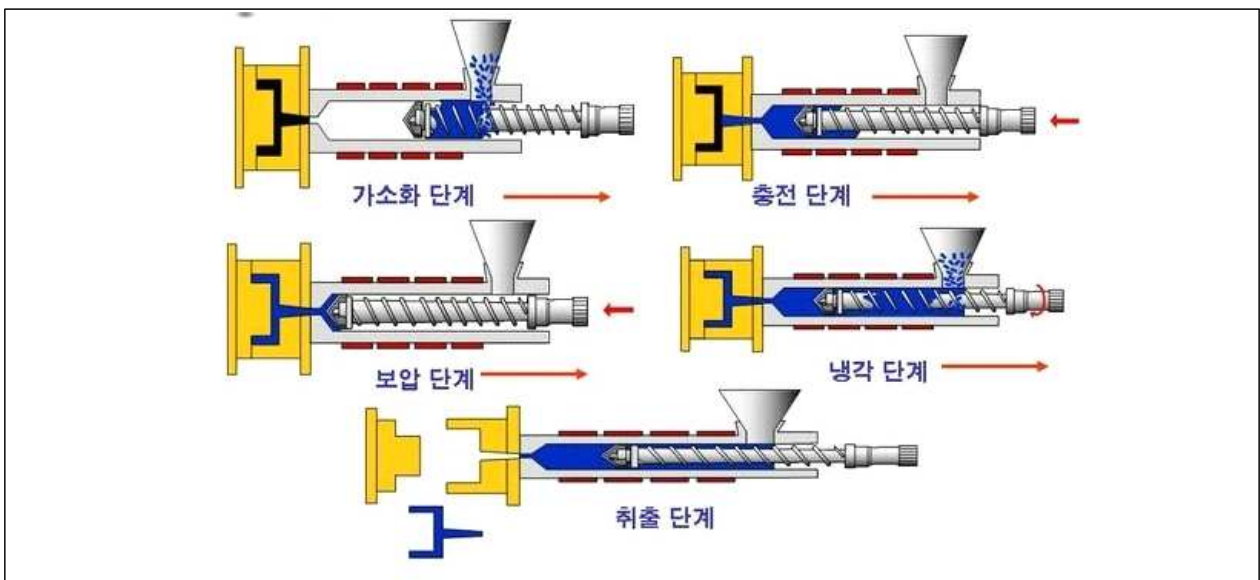
□ 사출 성형의 장점

- 형상이 복잡한 제품 생산에 적합
- 하나의 금형으로 여러 가지 소재의 제품을 생산할 수 있음(다양한 재료 및 색상의 제품 생산 가능)
- 다른 생산 기술에 비해 비용이 저렴
- 성형 사이클이 짧고 대량 생산 시 높은 생산 효율을 가짐
- 성형 공정의 자동화와 제어가 가능

□ 사출 성형의 단점

- 지정 금형 제작이 필요하여 초기 비용이 높음
- 금형 제조 후의 수정 비용이 많음

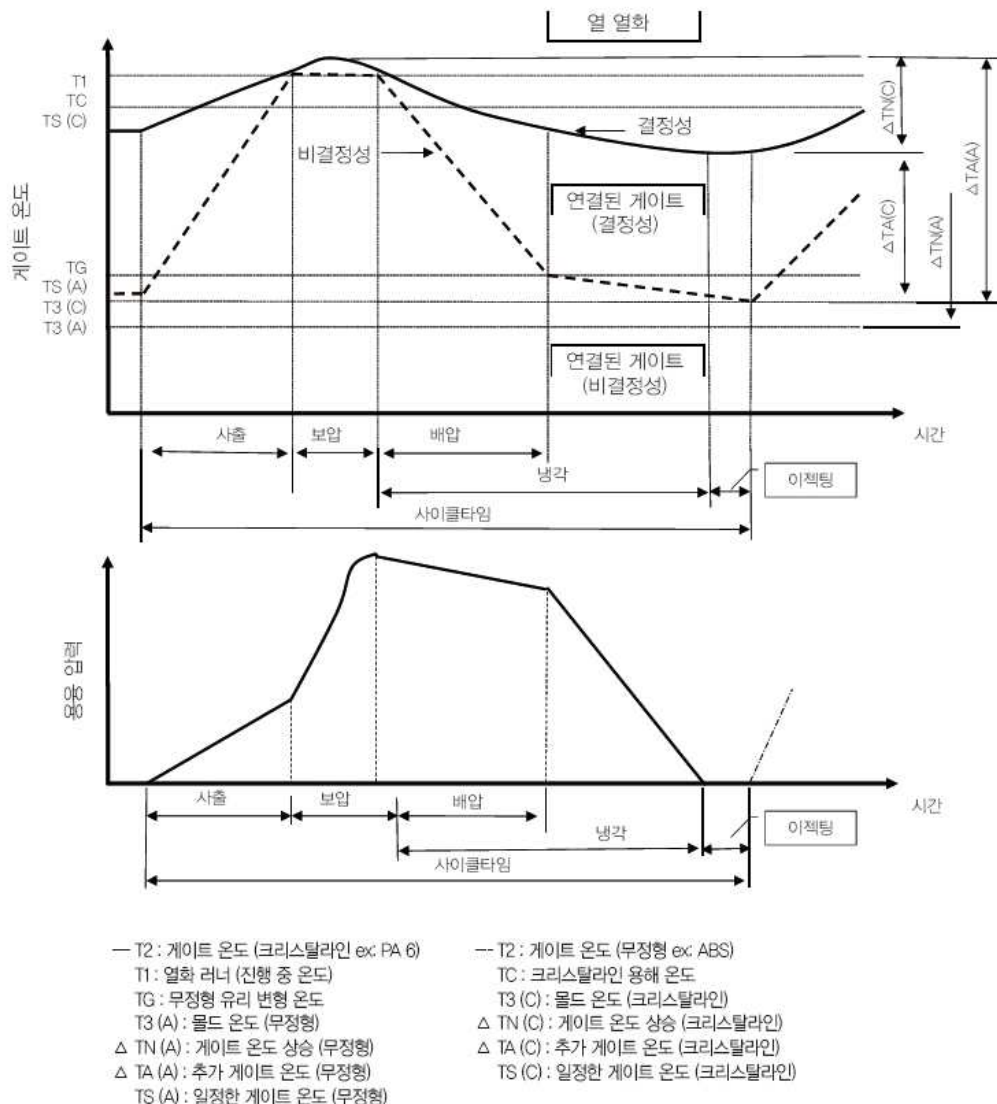
□ 사출 성형의 개념



[그림 1] 사출 성형 단계

- 사출 성형의 단계는 크게 구분하여 충전, 보압, 냉각, 취출 단계로 나눌 수 있다.
- ① 가소화 단계(Plasticization Phase) : 고체의 플라스틱 재료가 호퍼를 통해 회전하고 있는 스크류의 후부 날개 부분으로 이송되고 스크류가 회전하면서 후퇴하는 동안 실린더의 선단 부분으로 원료가 이송되는 동안 딱딱한 상태의 원료는 실린더를 에워싸고 있는 밴드 히터로부터 전도된 열과 원료에 작용 되는 전단력에 의해서 발생 되는 마찰열에 의해 유동성이 좋게 원재료를 가열하는 것을 말한다.

- ② 충전 단계(Filling Phase) : 스크류의 후단 부분에 부착된 유압 실린더의 힘으로 스크류가 회전하며 앞으로 나갈 때 실린더 선단 부분에 있는 용융된 원료는 금형의 캐비티 내부로 이동해 들어가 캐비티를 충전하게 되는데 이 단계를 말함.
- ③ 보압 단계(Compensating Phase) : 밀폐된 금형에 고온의 원재료가 투입되면 일부 팽창 압력이 내부에 발생하게 되며 이 압력에 의해 원재료가 역류하는 현상을 막기 위해 일정한 압력으로 버텨주는 단계로서 캐비티 내 85% 충진을 완료하고 부족한 부피를 일정하게 보충한다.
- ④ 냉각 단계(Cooling Phase) : 성형이 완료된 후 고온의 수지를 냉각시키기 위해 냉각매체를 공급하는 과정을 말함. 보압이 끝난 후 게이트가 응고하게 되고 스크류가 후진하기 시작하며 수지의 열이 금형을 통해 냉각 채널 내 냉각수로 전달되어 수지가 냉각된다. 냉각 단계에서는 생산성 향상을 위한 냉각효과 및 제품의 균일한 온도분포 유지 등의 사항을 고려해야 한다.
- ⑤ 취출 단계(Extraction Phase) : 사출기의 가동측에 설치되어 있는 이젝터 봉이 유압실린더의 작동에 의해 앞으로 전진하여 금형에 설치되어 있는 이젝터 봉을 앞으로 전진시켜 이젝터 판을 밀고 핀이 제품을 밀어내어 취출 시킨다.



(출처: 박균명 박사 - 한국생산기술연구원 금형기술센터)

[그림 2] 결정성 수지와 비결정 수지의 게이트 주변 온도 변화 성형 조건 도식화

□ 성형조건의 설정

- 성형품의 품질저하에 영향을 주는 성형조건에 있어서는 수지온도, 사출속도와 압력, 균일한 냉각에 의해서 형성된다. 이러한 관계에 있어서 문제가 야기되고 있는 것이며 성형품의 형상, 크기 사용 재료에 따라 성형성은 더욱 복잡하게 되어 각 성형품 별로 표준화를 작성하지 않으면 안된다.

일반적인 방법으로는

- 온도를 일정하게 하고 저압 (고압) 저속 (고속)
- 압력을 일정하게 하고 고온 (저온) 고속 (저속)
- 온도를 일정하게 하고 고온 (저온) 고압 (고압) 의 순으로 조합을 이루어 사출조건을 정한다.

① 실린더 온도

- 재료의 온도, 재료 GRADE, 금형구조, 사출 성형기에 따라 차이가 있으며 제품표면에 광택을 요구하는 성형품의 성형에는 실린더 온도를 높게 설정하면 안되고 실린더 온도를 높이면 충격강도가 높아지는 경우가 있으나 이러한 요인은 재료의 특성별로 차이가 있다. 극단적으로 온도를 높게 설정하면 재료의 열분해, FLASH가 발생하기 쉽다. 제품 두께가 두꺼운 제품은 유동저항이 적기 때문에 실린더 온도는 재료의 유동성과 크게 영향을 미친다.

② 사출 압력에 따른 사출속도

- 일반적으로 사출 속도와 압력은 초기 주입시에 높게 설정하여 정한다. 두꺼운 제품이 성형에 있어서는 속도를 빠르게 하는 것이 중요하다. CAVITY 내의 수지 온도를 균일하게 하여야 강도, 밀도를 균일하게 할 수 있다. 사출압력과 속도는 CAVITY FLOW ROUTE의 단면적에 관계가 있으며 제품설계 시 (금형 설계) GATE 위치에 유의해야 한다.

③ 금형 온도

- 금형 온도 설정도 일반적으로 높게 설정하며 이유는 유동저항을 적게 해주는 데 있으며 또한 잔류응력을 적게 해주는 데 있다. 금형 온도를 높여주면 광택의 효과를 얻을 수 있으며 CAVITY CORE의 금형 표면 온도를 일정하게 유지시켜 주는 것이 이상적이다. 금형 온도를 높여주면 성형수축이 크게 발생하는 경우가 있어 재료의 특성별로 조건설정이 중요하며 외관 제품은 금형 온도를 높게, 기능 및 치수가 중요한 제품은 금형 온도를 낮게 해주어야 한다.

④ 사출시간

- 성형품의 형상, 크기에 따라 차이가 있으며 완전 충전과 외관상의 문제를 결정하는 요인이 된다. 두꺼운 제품에는 SINK 또는 VOID의 발생을 방지해야되며 GATE의 폭이 짧거나 작거나 하는 경우에 따라 사출 압력, 재료의 온도와 상관관계가 있다. 사출시간이 짧으면 외관상으로 WRINLE MARK가 발생하며 VOID발생으로 인한 충격강도 저하의 요인이 될 수 있다.

- ⑤ 실린더온도, 금형온도, 스프루, 런너 및 성형품의 두께에 따라 크게 변한다. 충분한 냉각 시간으로서 성형품의 변형을 적게 해주면 성형 사이클이 길어지고 너무 짧게 하면 취출 후 경시 변화에 의한 평면 밀도를 이루면서 왜곡, 변형, 균열 등이 발생한다. 냉각 시간 설정은 수차례에 걸친 외관측정과 촌법측정으로서 결정되어야 한다.

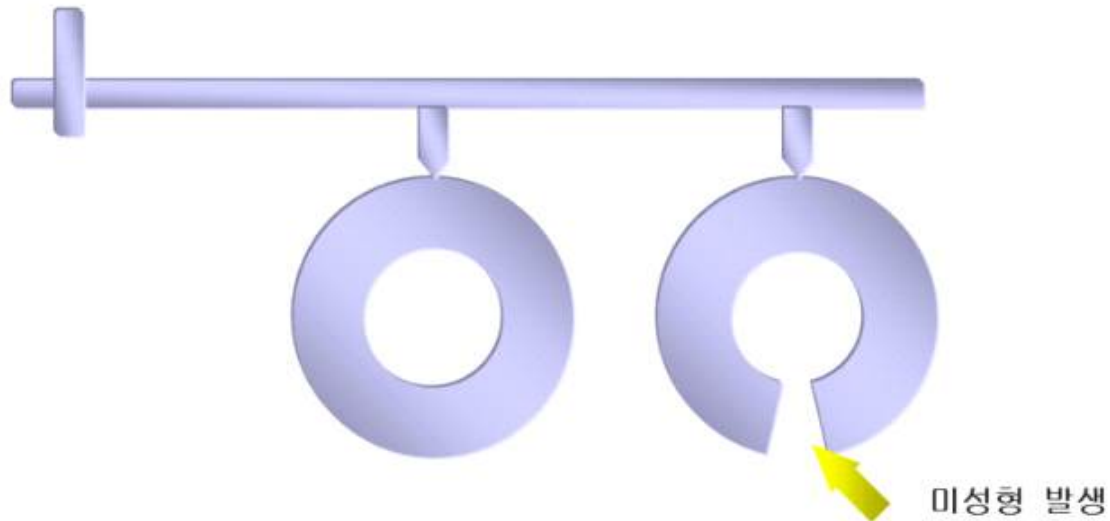
나. 이슈사항(Pain point)

□ 사출 성형에 있어서 성형 불량률의 원인은 크게 10가지로 구분할 수 있다.

- 충전 부족 (Short Shot) : 성형할 수지가 성형기 (plastic molding machine)의 실린더 안에서 충분히 가열되지 않거나 사출 압력과 금형 온도가 매우 낮을 경우, 금형 전체에 수지가 들어가지 않고 냉각 고화해서 성형품의 일부가 모자라는 현상이다.
- 주요 원인은 다음과 같다.

- ① 수지의 유동성이 부족하다.

- ② 금형내압이 부족하다.
- ③ 성형기의 능력이 부족하다.
- ④ cavity 안의 공기빠짐이 불량하다.
- ⑤ 재료공급량이 부적정하다.
- ⑥ 유동저항이 너무 크다.



[그림 3] 충전 부족(Short Shot)

2. 분석 목표

가. 분석 목표

□ 사출공정 주요변수 정의

- 사출공정에서 수집되는 다양한 변수들 중 사출품의 품질에 영향을 미치는 변수들을 정의
- 정의된 변수들을 토대로 현 기업의 데이터수집 및 관리현황 등을 파악하여 분석의 신뢰도 점검
- 정의된 변수들의 수집주기, 수집경로, 데이터 값 확인 등을 통하여 데이터 신뢰도 확보

□ 데이터 품질 전처리

- 실제 공정상 발생하는 데이터들 중 일부는 의미가 없는 값이나 누락 및 오타가 발생하여, AI모델 개발 시 모델의 품질에 부정적인 영향을 미칠 수 있음
- 좋은 AI모델이라고 하더라도 데이터 품질이 낮은 데이터를 활용하면 모델의 결과가 좋을 수 없음
- 따라서, 데이터 품질 전처리는 데이터 분석에 있어서 필수적인 단계이며, 데이터들의 품질 지수를 일차적으로 파악하고, 데이터 전처리를 통해 품질 지수를 향상시킬 필요가 있음

□ 데이터 품질 지수

- 완전성(Completeness): 필수항목에 누락이 없어야함
- 유일성(Uniqueness): 데이터 항목은 유일해야 하며 중복되어서는 안됨
- 유효성(Validity): 데이터 항목은 정해진 데이터 유효범위 및 도메인을 충족해야 함
- 일관성(Consistency): 데이터가 지켜야 할 구조, 값, 표현되는 형태가 일관되게 정의되고, 서로 일치해야함
- 정확성(Accuracy): 실제 존재하는 객체의 표현 값이 정확하게 반영이 되어야 함
- 무결성(Integrity): 데이터베이스 자료의 오류 없이 변화에 영향을 받지 않고 데이터의 유일성, 유효성, 일관성이 보호되어야 함

□ 품질 분류 AI 모델 개발

- 전처리 된 공정데이터들을 기반으로 품질 분류 AI 모델을 개발
- 사출공정에서 발생하는 공정 데이터와 품질 데이터 간의 상관관계 분석
- 정확한 품질 분류를 통해 품질 향상 및 비용 절감 가능할 것으로 기대

나. 데이터 정의 및 소개

□ 사출공정 데이터

- 사출기에서 수집되는 사출공정 데이터를 MongoDB에 저장한 데이터를 원천 데이터 셋으로 함
- 본 공정데이터를 AI모델의 input 변수로 활용되는 독립변수로 활용하며, 그 변수들의 특성에 따라 온도, 압력, 시간, 속도, 양 관련 변수들로 군집할 수 있음
- 온도 관련 변수: 스크류 온도, 실린더 온도, 수지 온도, 금형 온도, 건조 온도, 유압 온도 등
- 압력 관련 변수: 충전 압력, 보압, 배압, 이형 압력, 형개 압력, 형체 압력 등
- 시간 관련 변수: 충전 시간, 보압 시간, 냉각 시간, 건조 시간 등
- 속도 관련 변수: 사출 속도, 스크류 회전 속도, 형개 속도, 이형 속도 등
- 양 관련 변수: 계량, 이형량, 쿠션량 등

□ 사출품질 데이터

- MongoDB에 저장되는 tabular 형태의 사출공정 데이터에 개별적으로 매칭되는 학습용 사출품질 데이터
- 변수 명은 임의적으로 Passorfail 으로 지정되어있으며, 0일 경우 해당 생산품은 불량품, 1일 경우 해당 생산품은 양품을 의미함
- 본 데이터를 품질 분류 AI모델의 종속변수로 활용함

다. 데이터 분석 기대효과 및 시사점 요약 기술

□ 데이터 분석 기대효과

- 사출공정에서 수집된 사출관련 온도, 속도, 시간 등의 데이터를 사용하여 생산품의 품질을 판정하는 AI 모델을 만들고 주요 변수를 추출하여 원인 분석을 진행할 수 있음
- 이는 사람이 육안으로 검사하는 것보다 높은 검사 정확도로 사출공정의 제품 생산의 품질 향상과 노동력의 비용 절감이 가능함
- 또한, Feature importance 테스트를 통하여, AI 모델의 판정에 영향을 미치는 AI 관점의 주요 변수 추출을 통하여 품질에 영향을 주는 변수를 파악할 수 있음

□ 시사점(implication) 요약

- 대부분의 사출공정에서는 수집되는 공정 데이터를 활용하여 생산품의 품질을 판정하지 않고, 작업자의 주관적인 판단에 의존하여 제품의 품질을 판단함
- 하지만, 작업자의 오분류 등으로 인하여 품질 저하 등의 이슈가 있음
- 또한, 사람의 의한 판단으로 품질의 균일성을 확보하기 힘들었음
- 따라서, 본 가이드북에서는 사출공정에서 수집된 데이터를 바탕으로 품질분류 AI 분석 모델을 구축하고 해당 모델을 통해 생산품의 품질을 분류하고자 함
- 기존 공정에서는 데이터의 수집은 이루어지고 있었으나 분석이 진행되고 있지 않았음
- 본 분석은, 사출 공정의 품질 및 공정 데이터를 수집 후, 가공/전처리, AI 모델 개발과 제조 공정의 적용 및 검증을 통해 중소기업에 빅데이터 및 AI를 적용하여 실질적인 품질 및 비용 절감 개선에 기여 할 수 있다는 점에서 시사하는 바가 크다고 판단됨

1. 제조데이터 소개

가. 데이터 수집 방법

□ 제조분야

- 하우징, 임펠러, 커버 플레이트 등 자동차 부품 생산공정

□ 제조공정명

- 자동차 부품 사출공정

□ 수집장비

- 제조공정 내 사출기 1대에서 수집되는 공정 데이터
- 해당 생산품 품질 데이터

□ 수집기간

- 2022년 10월 3일 ~ 2022년 11월 6일

□ 수집주기

- 사출 Shot 단위 수집

나. 데이터 유형/구조

□ 데이터셋 구조

- MongoDB에 사출 shot 별 데이터들이 가공되어 수집됨
- 공정 데이터 및 품질 데이터가 tabular 포맷으로 저장되며 그 예시는 다음과 같음

Date	EQUIP_CD	WorkingNum	Passorfail	Reason	Injection_1	Filling_Tim	Plasticizing	Cycle_Tim	Clamp_Clc	Cushion_F	Plasticizing	Clamp_Op
2022-10-03	2	2103042011	0	1	13.18574	6.028869	15.16857	54.09298	6.595035	601.2055	63.12157	596.1508
2022-10-03	2	2103042012	0	1	15.18497	5.559804	17.58356	62.12707	7.608364	692.5844	72.71431	686.8694
2022-10-03	2	2103042013	0	1	13.93046	3.861301	16.25663	56.75825	6.938782	633.828	66.53576	628.5503
2022-10-03	2	2103042014	0	1	17.26926	7.760626	19.97997	71.88326	8.523193	784.2125	82.31717	777.588
2022-10-03	2	2103042015	0	1	13.67435	6.855769	16.02112	57.19088	6.893439	627.3829	65.84801	622.0704
2022-10-03	2	2103042016	0	1	14.75186	6.642863	17.15822	60.15837	7.41138	673.0704	70.64895	667.4297
2022-10-03	2	2103042017	0	1	16.4055	6.522334	19.09225	66.34547	8.167011	744.7973	78.18153	738.7086
2022-10-03	2	2103042018	0	1	13.66851	5.914269	15.93959	55.77394	6.844964	627.3979	65.86009	622.0704
2022-10-03	2	2103042019	0	1	14.40133	8.149911	16.78977	57.79944	7.124185	653.3343	68.60974	647.99
2022-10-03	2	2103042020	0	1	16.58643	7.046042	19.28101	67.25314	8.259587	758.1724	79.55083	751.6684
2022-10-03	2	2103042021	0	1	16.26455	6.695394	18.84077	66.3516	8.048834	738.3222	77.52556	732.2287
2022-10-03	2	2103042022	0	1	12.35907	6.085828	14.46566	48.93727	6.186781	562.0302	59.00522	557.2714
2022-10-03	2	2103042023	0	1	12.80314	4.263532	14.50137	52.93223	6.360446	581.504	61.05181	576.7111
2022-10-03	2	2103042024	1	0	8.700018	4.230167	15.20299	54.51294	6.405859	588.0686	62.01678	583.191

[그림 4] 수집 데이터셋 예시(1)

□ 데이터 개수

- 2022년 10월 3일부터 2022년 11월 6일까지 수집된 데이터로 총 11,280건의 생산 및 품질 데이터가 존재함

Cushion_F	Plasticizing	Clamp_Op	Max_Inject	Max_Screw	Average_S	Max_Inject	Max_Switch	Max_Back	Average_E
601.2055	63.12157	596.1508	44.84173	28.60689	269.0922	133.5989	133.8677	58.43249	52.19307
692.5844	72.71431	686.8694	44.63049	32.22445	309.9928	151.4247	151.5758	71.11635	52.05384
633.828	66.53576	628.5503	45.6397	29.89682	283.6725	142.081	135.0637	40.69876	75.9961
784.2125	82.31717	777.588	60.82845	37.27668	350.683	169.3228	167.9427	45.16125	90.31881
627.3829	65.84801	622.0704	37.10785	29.60992	280.6944	140.1045	142.753	37.60412	54.3618
673.0704	70.64895	667.4297	44.67129	31.77711	301.309	149.799	149.4587	41.75923	86.6298
744.7973	78.18153	738.7086	56.94747	34.89167	333.3059	161.802	161.4528	12.40628	79.74331
627.3979	65.86009	622.0704	41.05242	29.84413	280.7587	140.4609	137.2394	46.6182	81.02873
653.3343	68.60974	647.99	37.32681	31.06588	292.4436	146.7841	138.9844	70.1092	78.83019
758.1724	79.55083	751.6684	66.27176	35.74665	339.2667	165.5088	175.0277	51.46593	65.87501
738.3222	77.52556	732.2287	49.39495	35.01424	330.3493	163.3228	160.9716	60.67636	62.72891
562.0302	59.00522	557.2714	32.98254	26.51477	251.5473	126.8803	117.6674	54.67418	56.28039
581.504	61.05181	576.7111	38.3668	27.27146	260.2477	125.0535	131.388	42.49087	57.01914
588.0686	62.01678	583.191	51.53351	27.69587	263.247	128.1263	123.8364	33.7955	54.4404

[그림 5] 수집 데이터셋 예시 (2)

□ 데이터셋 정의

- 사출기에서 수집되는 데이터는 아래와 같이 정의됨

변수명	변수설명	단위
Passorfail	사출품 품질 판정결과 (0=불량품, 1=양품)	-
TimeStamp	시간(YY:HH:MM:SS)	-
PART_FACT_PLAN_DATE	생산 지시일	-
PART_FACT_SERIAL	생산 코드	-
Shot_Number	Shot 수	-
PART_NO	제품의 모델코드	-
PART_NAME	제품의 이름	-
EQUIP_CD	생산한 사출기 호기	-
EQUIP_NAME	생산한 사출기 모델명	-
PART_FACT_PLAN_QTY	해당 품목에 대한 생산지시 수량	수량 (개)
ERR_FACT_QTY	생산시 발생한 불량률의 개수	수량 (개)
FACT_QTY	생산시 발생한 합격률의 개수	수량 (개)
Injection_Time	고압+사출시간(고압(사출압) : 재료를 금형에 유입시킬때의 압력), 사출시간 : 재료를 금형에 유입시키는데 소요되는 시간)	Sec (초)
Filling_Time	충진시간으로 사출기에서 금형으로 내용물이 주입되는 시간	Sec (초)
Plasticizing_Time	계량시간으로 재료를 스크류에 1번 생산할 만큼 용융되어 저장되는 시간	Sec (초)
Cycle_Time	1번의 제품생산에 소요되는 생산시간	Sec (초)
Clamp_Close_Time	제품이 생산되고 난후 열려있는 금형을 사출기가 닫아주고 빈틈이 없이 고정축과 이동축을 꼭 잡아주는데 걸리는 시간	Sec (초)
Cushion_Position	보압(사출압의 다음으로 가해지는 압력(금형내부압력을 조절하여 과충전을 방지))을 하기위한 스크류의 위치	mm
Switch_Over_Position	고압,보압절환위치(고압(사출압)에서 보압으로 진행될때의 위치))	mm
Plasticizing_Position	계량완료위치(계량을 마친 스크류의 위치)	mm

Clamp_Open_Position	제품이 생산되어 추출하기위해 금형이 열리고 난 위치	mm/s
Max_Injection_Speed	배럴에 계량되어 있는 용융수지가 금형으로 흘러들어가는데 측정되는 최대속도	mm/s
Max_Screw_RPM	사출을 위한 스크류의 최대속도	mm/s
Average_Screw_RPM	사출을 위한 스크류의 평균속도	mm/s
Max_Injection_Pressure	배럴에 계량되어 있는 용융수지가 금형으로 흘러들어가는데 가해지는 최대압력	MPa
Max_Switch_Over_Pressure	사출에서 보압(충진된 수지가 밀리지않게 압력을준다)으로 변환되는 압력	MPa
Max_Back_Pressure	수지가 계량이 되는중에 스크류가 밀려나는 현상을 저지하기위한 최대압력	MPa
Average_Back_Pressure	수지가 계량이 되는중에 스크류가 밀려나는 현상을 저지하기위한 평균압력	MPa
Barrel_Temperature_1	계량 및 사출시 수지가 일정하게 용융(녹임)을 유지하기위해 온도가 일정해야한다	섭씨 (°C)
Barrel_Temperature_2		섭씨 (°C)
Barrel_Temperature_3		섭씨 (°C)
Barrel_Temperature_4		섭씨 (°C)
Barrel_Temperature_5		섭씨 (°C)
Barrel_Temperature_6		섭씨 (°C)
Barrel_Temperature_7		섭씨 (°C)
Hopper_Temperature	재료주입구의 온도(충분히 건조시켜주며 재료가 용융되는시간을 절약시켜주기위해 온도가 높아야한다)	섭씨 (°C)
Mold_Temperature_1	재료주입구의 온도(충분히 건조시켜주며 재료가 용융되는시간을 절약시켜주기위해 온도가 높아야한다)	섭씨 (°C)
Mold_Temperature_2		섭씨 (°C)
Mold_Temperature_3		섭씨 (°C)
Mold_Temperature_4		섭씨 (°C)
Mold_Temperature_5		섭씨 (°C)
Mold_Temperature_6		섭씨 (°C)
Mold_Temperature_7		섭씨 (°C)
Mold_Temperature_8		섭씨 (°C)
Mold_Temperature_9		섭씨 (°C)
Mold_Temperature_10		섭씨 (°C)
Mold_Temperature_11		섭씨 (°C)
Mold_Temperature_12		섭씨 (°C)
Clamp_open_time	형개시간(사출이 끝나고 제품을 추출하기위해 금형이 열릴때 형개위치까지 열리는데 걸리는 시간)	Sec (초)

□ 주요 변수 기술 통계

- 사출공정 데이터 및 품질 데이터에 대해 기초통계분석을 진행할 수 있음
- 다양한 기초통계 중, 평균, 표준편차, 최소값 및 최대값을 확인하였음
- 기술 통계 계산 결과는 아래 표와 같음

구분	평균	표준편차	최소값	최대값
PassOrFail	0.94	0.11	0.00	1.00
Injection_Time	8.11	0.13	7.88	9.11
Filling_Time	3.97	0.12	3.11	4.27
Plasticizing_Time	16.82	0.27	16.32	21.10

Cycle_Time	42.11	0.33	36.94	47.24
Clamp_Close_Time	7.14	0.02	6.67	7.21
Cushion_Position	658.41	0.11	653.37	661.10
Switch_Over_Position	0.00	0.00	0.00	0.00
Plasticizing_Position	64.28	0.43	59.76	68.86
Clamp_Open_Position	646.31	43.69	69.64	647.99
Max_Injection_Speed	55.22	3.49	38.50	64.70
Max_Screw_RPM	30.70	0.24	30.30	31.20
Average_Screw_RPM	111.31	131.11	29.20	293.90
Max_Injection_Pressure	142.10	1.72	140.70	169.10
Max_Switch_Over_Pressure	136.72	0.89	122.40	143.70
Max_Back_Pressure	38.07	1.66	27.70	85.20
Average_Back_Pressure	59.50	3.75	29.97	95.20
Barrel_Temperature_1	266.03	2.31	254.90	277.90
Barrel_Temperature_2	265.17	2.11	259.60	276.60
Barrel_Temperature_3	264.91	1.65	259.60	276.00
Barrel_Temperature_4	260.32	1.67	244.70	272.40
Barrel_Temperature_5	264.94	0.31	269.90	276.40
Barrel_Temperature_6	269.98	0.36	264.60	270.70
Barrel_Temperature_7	0.00	0.00	0.00	0.00
Hopper_Temperature	67.118	4.27	38.50	70.60
Mold_Temperature_1	0.00	0.00	0.00	0.00
Mold_Temperature_2	0.00	0.00	0.00	0.00
Mold_Temperature_3	22.299	2.98	18.80	25.30
Mold_Temperature_4	23.557	0.99	19.80	27.80
Mold_Temperature_5	0.00	0.00	0.00	0.00
Mold_Temperature_6	0.00	0.00	0.00	0.00
Mold_Temperature_7	0.00	0.00	0.00	0.00
Mold_Temperature_8	0.00	0.00	0.00	0.00
Mold_Temperature_9	0.00	0.00	0.00	0.00
Mold_Temperature_10	0.00	0.00	0.00	0.00
Mold_Temperature_11	0.00	0.00	0.00	0.00
Mold_Temperature_12	0.00	0.00	0.00	0.00

□ 독립변수 정의

- 본 가이드북에 사용된 독립변수들은 사출기에서 수집된 데이터로, 독립변수들의 리스트는 다음과 같음

독립변수명	변수설명	단위
TimeStamp	시간(YY:HH:MM:SS)	-
PART_FACT_PLAN_DATE	생산 지시일	-
PART_FACT_SERIAL	생산 코드	-
Shot_Number	Shot 수	-
PART_NO	제품의 모델코드	-
PART_NAME	제품의 이름	-
EQUIP_CD	생산한 사출기 호기	-

EQUIP_NAME	생산한 사출기 모델명	-
PART_FACT_PLAN_QTY	해당 품목에 대한 생산지시 수량	수량 (개)
ERR_FACT_QTY	생산시 발생한 불량률의 개수	수량 (개)
FACT_QTY	생산시 발생한 합격률의 개수	수량 (개)
Injection_Time	고압+사출시간(고압(사출압) : 재료를 금형에 유입시킬때의 압력), 사출시간 : 재료를 금형에 유입시키는데 소요되는 시간)	Sec (초)
Filling_Time	충진시간으로 사출기에서 금형으로 내용물이 주입되는 시간	Sec (초)
Plasticizing_Time	계량시간으로 재료를 스크류에 1번 생산할 만큼 용융되어 저장되는 시간	Sec (초)
Cycle_Time	1번의 제품생산에 소요되는 생산시간	Sec (초)
Clamp_Close_Time	제품이 생산되고 난후 열려있는 금형을 사출기가 닫아주고 빈틈이 없이 고정축과 이동축을 꼭 잡아주는데 걸리는 시간	Sec (초)
Cushion_Position	보압(사출압의 다음으로 가해지는 압력(금형내부압력을 조절하여 과충전을 방지))을 하기위한 스크류의 위치	mm
Switch_Over_Position	고압,보압절환위치(고압(사출압)에서 보압으로 진행될때의 위치))	mm
Plasticizing_Position	계량완료위치(계량을 마친 스크류의 위치)	mm
Clamp_Open_Position	제품이 생산되어 추출하기위해 금형이 열리고 난 위치	mm/s
Max_Injection_Speed	배럴에 계량되어 있는 용융수지가 금형으로 흘러들어가는데 측정되는 최대속도	mm/s
Max_Screw_RPM	사출을 위한 스크류의 최대속도	mm/s
Average_Screw_RPM	사출을 위한 스크류의 평균속도	mm/s
Max_Injection_Pressure	배럴에 계량되어 있는 용융수지가 금형으로 흘러들어가는데 가해지는 최대압력	MPa
Max_Switch_Over_Pressure	사출에서 보압(충진된 수지가 밀리지않게 압력을준다)으로 변환되는 압력	MPa
Max_Back_Pressure	수지가 계량이 되는중에 스크류가 밀려나는 현상을 저지하기위한 최대압력	MPa
Average_Back_Pressure	수지가 계량이 되는중에 스크류가 밀려나는 현상을 저지하기위한 평균압력	MPa
Barrel_Temperature_1	계량 및 사출시 수지가 일정하게 용융(녹임)을 유지하기위해 온도가 일정해야한다	섭씨 (°C)
Barrel_Temperature_2		섭씨 (°C)
Barrel_Temperature_3		섭씨 (°C)
Barrel_Temperature_4		섭씨 (°C)
Barrel_Temperature_5		섭씨 (°C)
Barrel_Temperature_6		섭씨 (°C)
Barrel_Temperature_7		섭씨 (°C)
Hopper_Temperature	재료주입구의 온도(충분히 건조시켜주며 재료가 용융되는시간을 절약시켜주기위해 온도가 높아야한다)	섭씨 (°C)
Mold_Temperature_1	재료주입구의 온도(충분히 건조시켜주며 재료가 용융되는시간을 절약시켜주기위해 온도가 높아야한다)	섭씨 (°C)
Mold_Temperature_2		섭씨 (°C)
Mold_Temperature_3		섭씨 (°C)
Mold_Temperature_4		섭씨 (°C)
Mold_Temperature_5		섭씨 (°C)
Mold_Temperature_6		섭씨 (°C)
Mold_Temperature_7		섭씨 (°C)

Mold_Temperature_8		섭씨 (°C)
Mold_Temperature_9		섭씨 (°C)
Mold_Temperature_10		섭씨 (°C)
Mold_Temperature_11		섭씨 (°C)
Mold_Temperature_12		섭씨 (°C)
Clamp_open_time	형개시간(사출이 끝나고 제품을 추출하기 위해 금형이 열릴때 형개위치까지 열리는데 걸리는 시간)	Sec (초)

□ 종속변수 정의

- 본 가이드북에 사용된 종속변수는 품질 분류 결과이며, 그 데이터는 아래와 같음

종속변수명	변수설명	단위
Passorfail	사출품 품질 판정결과 (0=불량품, 1=양품)	-

다. 데이터 품질 전처리 및 품질 지표

□ 데이터 품질 지표

- 완전성, 유일성, 유효성, 일관성, 정확성, 무결성 등 6가지의 지표로 나누어 계산하여 데이터의 품질을 지표화함

□ 데이터 품질 계산 방법론

- 완전성 품질 지수 = $(1 - (\text{결측치} / \text{전체 데이터 수})) * 100$
 - 결측치 값이 30% 이상인 데이터들은 데이터의 완전성이 떨어지기 때문에 열별 결측치 값의 비율을 확인하여 삭제함
 - 데이터의 결측치를 확인하기 위해 'isnull()' 함수를 사용한 뒤 'sum()' 함수를 이용하여 총 결측치 개수를 구함
 - 구한 결측치의 개수를 이용하여 완전성 품질 지수를 구함
- 유일성 품질 지수 = $((\text{유일한 데이터 수} / \text{전체 데이터 수})) * 100$
 - 데이터 항목은 유일해야 하며 중복되어서는 안됨
 - 유일성을 판단하고 싶다면 날짜와 시간 두 열을 이용하여 합성키를 만든 뒤 기본키로 설정하여 중복을 판단
- 유효성 품질 지수 = $(\text{유효성 만족 데이터 수} / \text{전체 데이터 수}) * 100$
 - 본 데이터가 데이터셋에 정의된 수집 범위를 이용하여 조건(수집된 날짜, 데이터 형식)을 모두 충족하는지 확인
ex) 수집된 날짜 안에 들어가 있는가? 데이터가 형식에 맞는가?
- 일관성 품질 지수 = $(\text{일관성 만족 데이터 수} / \text{전체 데이터 수}) * 100$
 - 본 데이터가 구조, 값, 형태를 일관성 있게 지키고 있는지 확인함
 - 'date' 열의 데이터가 일관된 날짜 형태의 데이터를 가졌는지 유효성을 확인함
 - 'dtype()' 메서드를 이용하여 이외의 데이터 값들의 형태가 int 또는 float 타입인지 확인함
- 정확성 품질 지수 = $(1 - (\text{정확성 위배 데이터 수} / \text{전체 데이터 수})) * 100$
 - 실제 존재하는 객체의 표현 값이 정확히 반영되어야 한다. 본 데이터에서 칼럼값은 독립적이므로 정확성 지수를 확인하지 않음
- 무결성 품질 지수 = $(1 - (\text{유일성, 유효성, 일관성 지수 중 100%가 아닌 지수 개수} / 3)) * 100$
 - 무결성 품질 지수는 고려되는 모든 데이터가 유일성, 유효성, 일관성 지수를 만족하는지 확인하는 지수임. 본 데이터에서 유일성, 유효성, 일관성 모두 100%로, 세 가지 지수 모두 100%를 만족하므로 무결성 품질 지수는 100%

2. 분석 모델 소개

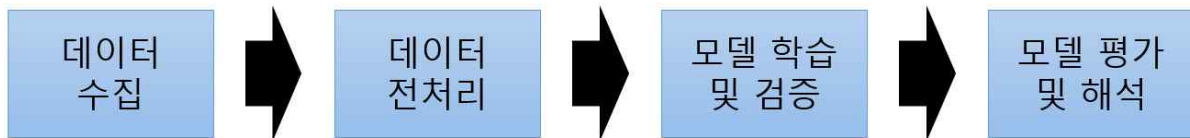
가. 데이터 흐름 및 인공지능 모델 적용 흐름도

□ 데이터 흐름

- 사출기에서 파생된 데이터는 수집 서버 내 MongoDB에 저장됨
- DB에 수집된 데이터의 품질 전처리 과정을 통해 학습용 데이터를 정제함
- 정제된 데이터를 기반으로 모델을 학습 및 평가함

□ AI 모델 적용 흐름도

- AI 모델의 적용 흐름도는 다음과 같음



[그림 5] AI 모델 적용 흐름도

나. AI 분석 모델

□ 사용 모델

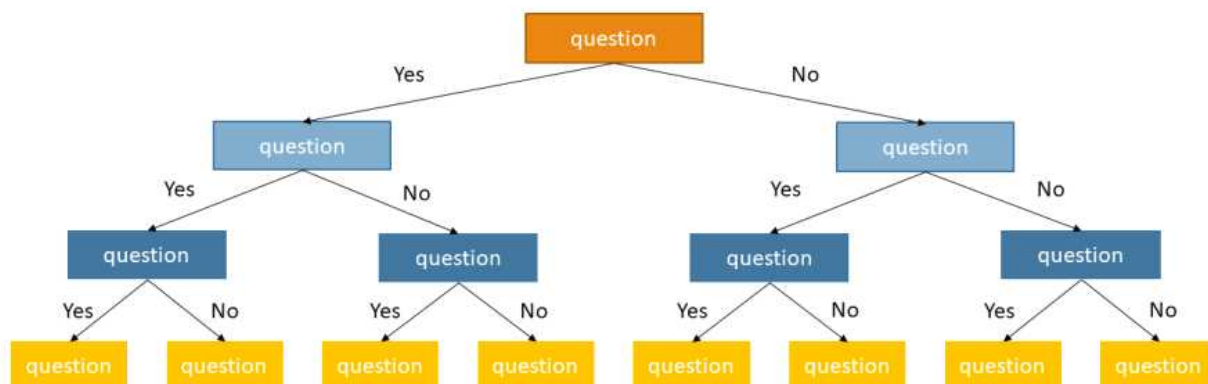
- Decision Tree, Random Forest 및 Adaboost

□ 선정 이유

- 위 알고리즘들은 분류 및 회귀 모두에 사용할 수 있는 트리기반 모델로 모델 생성에 영향도가 높은 주요 변수를 확인할 수 있고 이상치와 노이즈에 큰 영향을 받지 않으며, 종속변수가 범주형, 연속형 모두에서 사용 가능함
- 또한, 트리 시각화를 통하여 모델이 양품과 불량을 분류하는 각 변수의 기준을 확인할 수 있음
- 따라서 트리기반 모델은 빠르고 성능이 높으며, 모델의 결과를 해석할 수 있다는 장점이 있다.
- 머신러닝의 기본 모델이라고 할 수 있는 Decision Tree부터 Decision Tree의 단점인 과적합과 정확도를 개선 (앙상블, 부스팅)하여 많이 활용되고 있는 모델(AdaBoost)을 사용함으로써 모델별 차이와 다양한 모델에 대한 숙련도를 높이하고자 선정하였음

□ Decision Tree 모델 구체적 소개

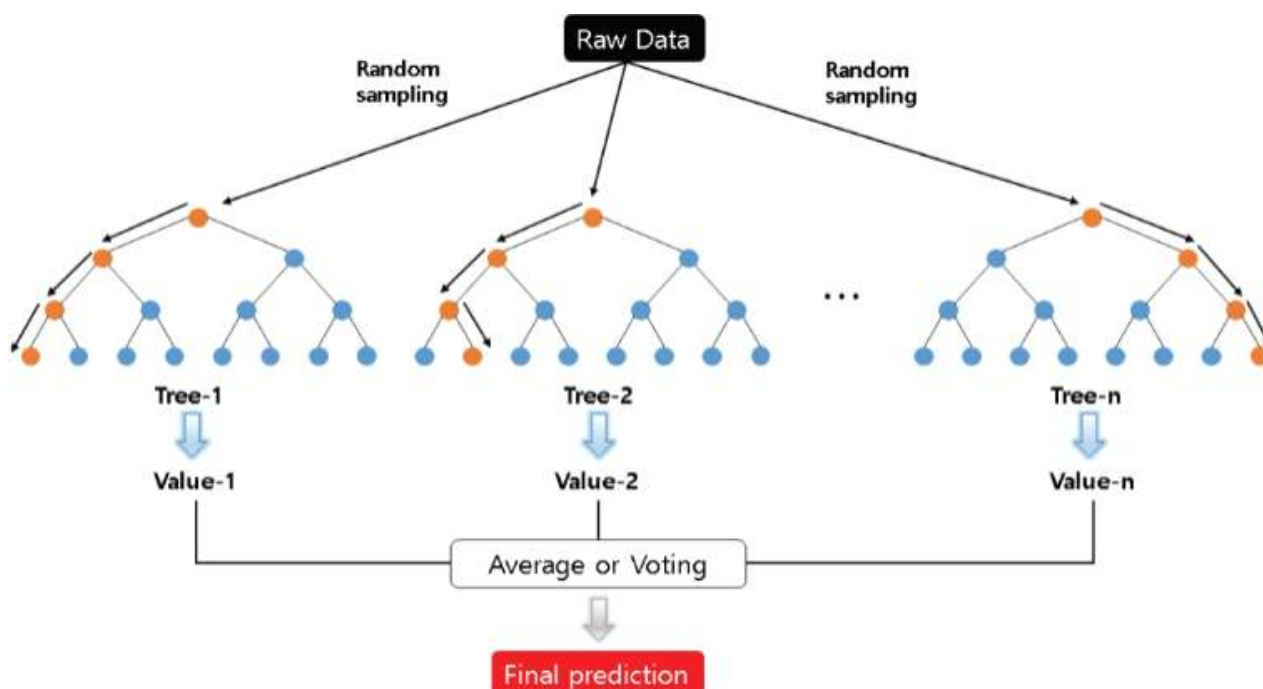
- Decision Tree는 특정 기준을 통해 데이터를 구분하는 지도 학습 모델로서 분류와 회귀 모두에 사용 가능함
- 한 줄기 마다 두 개의 노드를 생성하며 최초의 질문을 루트 노드(Root Node), 더 이상 분기하지 않는 경우를 터미널 노드 (Terminal Node) 또는 리프노드(Leaf Node)라고 함
- 구조가 나무를 뒤집어 놓은 것과 유사하여 Decision Tree라고 부르며, 이해하고 해석하기 쉽다는 장점이 있으나, 과적합 가능성이 높다는 단점이 있음. 이를 해결하기 위해 가지치기나 깊이를 조절함



[그림 6] Decision Tree 구조 예시 (출처: 김영미, 2020)

□ Random Forest 모델 구체적 소개

- Random Forest는 여러 학습 결과를 합치는 앙상블 방법 중 하나로 랜덤으로 복원 추출한 샘플로 다수의 decision tree를 생성한 후 각 tree가 예측한 값 중 가장 많은 값을 최종 예측 값으로 분류하는 알고리즘임
- 이를 통해 하나의 모델을 사용했을 때 발생하는 과적합 문제를 해결할 수 있음
- 결측치를 다루기 쉬우나 트리의 개수가 많아지면 시간이 오래 걸리는 단점이 있음

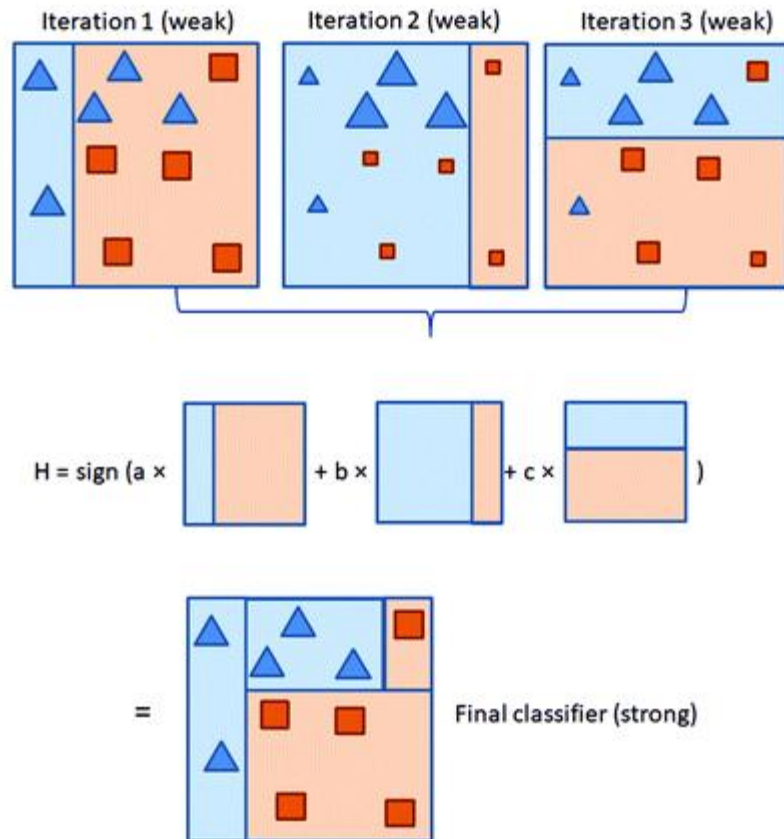


[그림 7] Random Forest 구조 예시 (출처: 김종성 외 5인, 2019)

□ Adaboost 모델 구체적 소개

- Adaboost는 약한 분류기(weak classifier)들이 상호보완 하도록 순차적(sequential)으로 학습하고, 이들을 조합하여 최종적으로 강한 분류기(strong classifier)의 성능을 향상시킬 수 있으며, 학습 데이터에 과적합 (overfitting) 현상이 적게 발생하여 예측 성능을 저하하지 않는 장점이 있음
- 약한 분류기(weak classifier)들은 한 번에 하나씩 순차적으로 학습을 진행함. 먼저 학습된 분류기는 제대로 분류를 해내는 데이터와 제대로 분류해내지 못하는 데이터들이 발생함

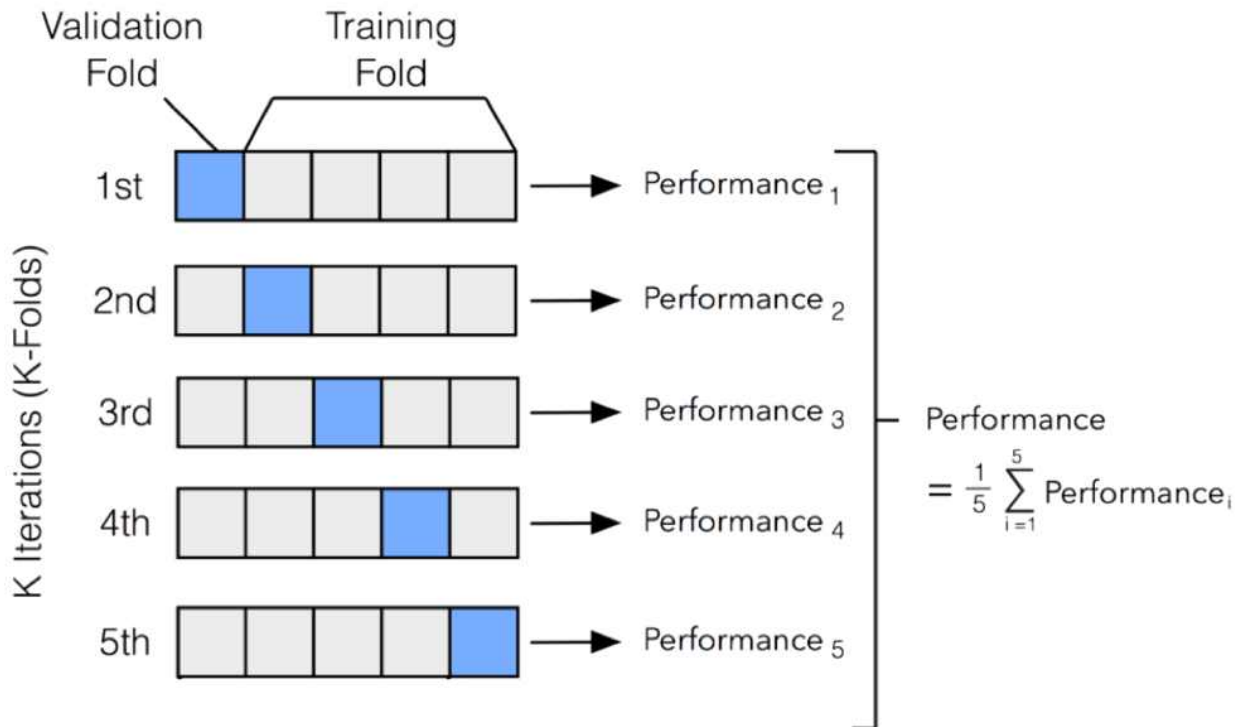
- 먼저 학습된 분류기가 제대로 분류한 결과 정보와 잘못 분류한 결과 정보를 다음 분류기에 전달함. 다음 분류기는 이전 분류기로부터 받은 정보를 활용하여 잘 분류해내지 못한 데이터들의 가중치(weight)를 높임
- 즉, 이전 분류기가 잘못 분류한 샘플의 가중치를 adaptive하게 바꿔가며 잘 못 분류되는 데이터에 더 집중하여 학습이 더 잘되게 함
- 이러한 특징 때문에 adaptive라는 이름이 붙게 되었고, 최종 분류기(strong classifier)는 이전에 학습한 약한 분류기들에 각각 가중치를 적용하고 조합하여 학습을 진행



[그림 8] Adaboost 구조 예시 (출처: Jiaqi Tang 외 2인, 2021)

□ 학습 및 검증 절차

- KFold 교차검증 방법론을 통하여 검증을 진행함
- KFold 교차검증은 데이터를 k개로 분할한 뒤, k-1개를 학습용 데이터 세트로, 1개를 평가용 데이터셋으로 사용하여 성능을 평가하는 방법임
- 데이터를 학습용과 평가용 데이터셋으로 여러 번 나눈 것의 평균적인 성능을 계산하면, 한 번 나누어서 학습하는 것에 비해 일반화된 성능을 얻을 수 있음



(출처: http://ethen8181.github.io/machine-learning/model_selection/model_selection.html)

[그림 9] Kfold 교차검증 설명 예시

□ AI 모델 구축 절차 설명

- 독립변수 및 종속변수 선정: 사출공정 데이터셋에서 Passorfail을 종속변수로 나머지 정제된 생산공정 데이터셋을 독립변수로 선정함
- 결측치 제거
 - Index 제거 방법 (dropna())을 통해 데이터 내 결측치가 존재할 경우 해당 행을 없애는 방법으로 결측치 제거를 Python에서 진행함
- 이상치 처리
 - 백분위 상한, 하한 0.1퍼센트 값을 이상치로 정의함
 - 해당 이상치를 모두 백분위 상한, 하한 0.1퍼센트 값으로 대체
- 학습(Train), 평가(Test) 데이터 분할: 모델을 평가할 때 학습 데이터를 사용하는 경우 실제 모델의 성능보다 우수한 결과가 나옴. 이를 방지하기 위해 학습 데이터와 평가 데이터로 분할하여 사용함
- 모델 학습: Decision Tree, Random Forest 및 Adaboost 모델들을 활용하여 사출공정에서 생산된 사출품의 품질을 분류하는 모델을 학습시킴
- 모델 검증: KFold 교차검증을 활용해 별도의 여러 세트로 구성된 학습 데이터 세트와 검증 데이터 세트에서 학습 및 검증을 수행함
- 모델 평가 및 해석: 혼동행렬(Confusion matrix)를 통해 모델의 성능을 평가하며, 정확도 (Accuracy) 와 F1-Score를 평가지표로 사용함

IV 분석 결과

1. 제조 데이터셋 수집 및 분석 결과

가. 제조 데이터셋 수집 내역

문제정의	필요 데이터	데이터셋 수집 내역
사출품 품질검사에 AI적용	- 사출품의 품질에 영향을 주는 주요 사출공정 변수들 - 학습용 사출품 품질 데이터	2022년 10월 3일 ~ 2022년 11월 4일 5주간 사출 공정 데이터 및 품질 데이터 수집 완료

나. 제조 데이터셋을 활용한 분석 결과

□ 학습(Train) 데이터 및 평가(Test) 데이터 분리

- 학습 데이터, 평가 데이터의 비율은 7:3으로 진행함
- 이때, Random_state는 임의로 42로 설정함. 이는 코드 실행마다 동일한 결과를 얻기 위해 지정하는 parameter임

□ 데이터 스케일링

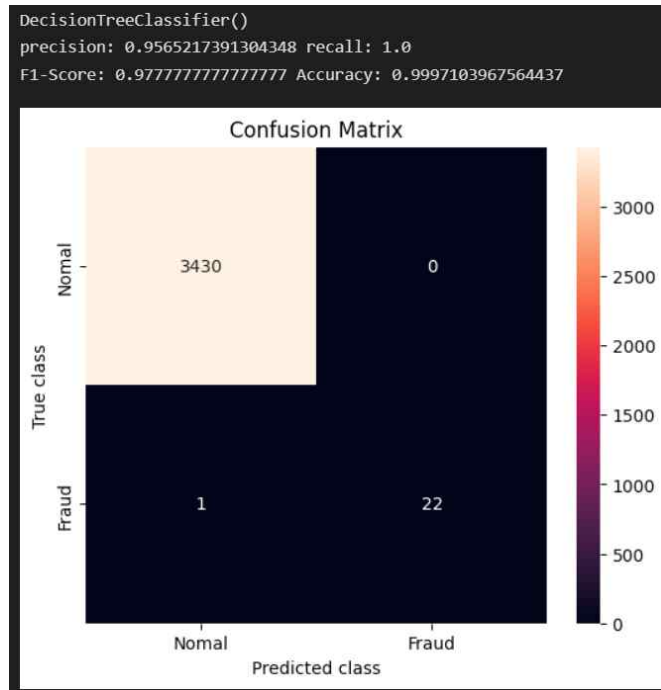
- 데이터 변수별로 데이터가 가진 크기와 편차가 다르기 때문에 한 특징을 과하게 반영하거나 패턴을 찾는 데 어려움이 생길 수 있음
- 훈련 데이터에 MinMaxScaler를 적용하여 모든 특성의 범위를 같게 만들어 줌
- MinMaxScaler 메서드는 변수의 범위를 조정하는 정규화 함수로, 모든 데이터가 0과 1 사이의 값을 갖도록 해주는 함수임

□ AI 모델링 진행

- sklearn 패키지의 'DecisionTreeClassifier()' 메서드, 'RandomForestClassifier()' 메서드, 'AdaBoostClassifier()' 메서드를 통해 Decision Tree, Random Forest, Adaboost 등 3가지의 트리 분류기 모델을 생성하여 학습 및 검증을 진행함

□ Decision Tree 모델 결과

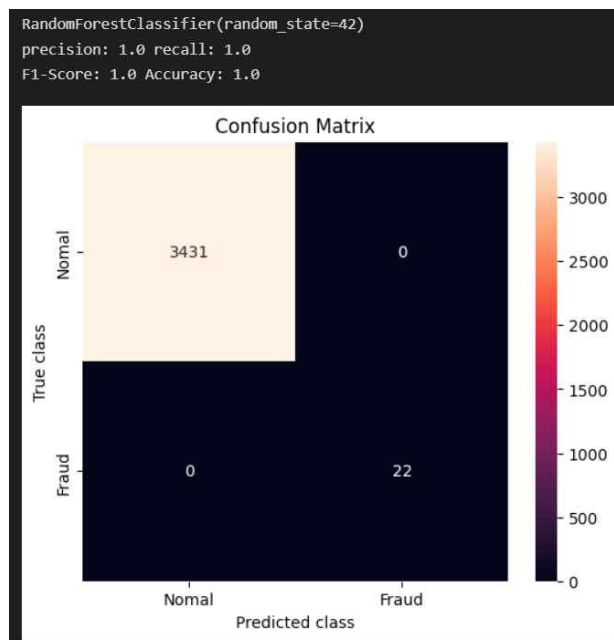
- KFold 교차 검증을 활용하여 데이터를 학습용과 평가용 데이터셋으로 여러 번 나누어 평균적인 성능을 계산하면, 한 번 나누어서 학습하는 것에 비해 일반화된 성능을 얻을 수 있음
- Decision Tree 모델을 학습데이터로 학습한 후 KFold 교차검증을 통해 검증을 진행함
- Fold 중 정확도가 가장 높은 Fold를 최적 모델로 선정함
- 선정된 최적의 모델을 활용하여 평가 데이터에 대해 평가를 진행하며, 그 결과인 혼동행렬은 아래와 같음
- 혼동행렬을 통해 확인할 수 있듯이, 총 3,453건의 데이터 중 3,452건에 대해서 정교하게 분류하였으며, 99.9%의 정확도를 확인할 수 있음



[그림 10] Decision Tree 모델 평가 결과

□ Random Forest 모델 결과

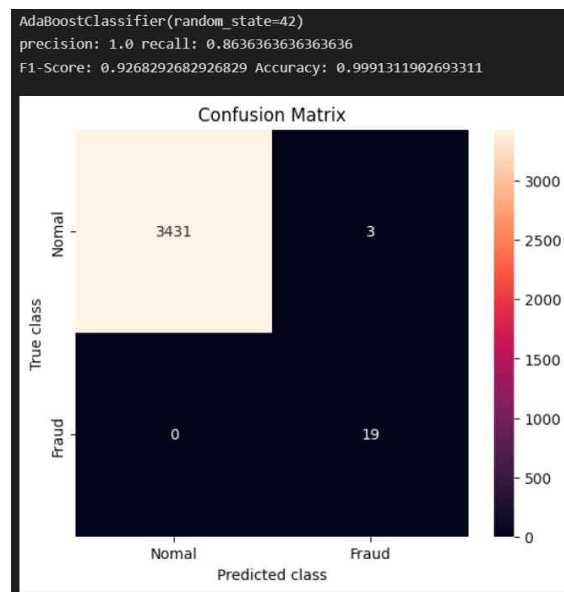
- KFold 교차 검증을 활용하여 데이터를 학습용과 평가용 데이터셋으로 여러 번 나누어 평균적인 성능을 계산하면, 한 번 나누어서 학습하는 것에 비해 일반화된 성능을 얻을 수 있음
- Random Forest 모델을 학습데이터로 학습한 후 KFold 교차검증을 통해 검증을 진행함
- Fold 중 정확도가 가장 높은 Fold를 최적 모델로 선정함
- 선정된 최적의 모델을 활용하여 평가 데이터에 대해 평가를 진행하며, 그 결과인 혼동행렬은 아래와 같음
- 혼동행렬을 통해 확인할 수 있듯이, 총 3,453건의 데이터 중 3,453건 모두에 대해서 정교하게 분류하였으며, 100%의 정확도를 확인할 수 있음



[그림 11] Random Forest 모델 평가 결과

□ Adaboost 모델 결과

- KFold 교차 검증을 활용하여 데이터를 학습용과 평가용 데이터셋으로 여러 번 나누어 평균적인 성능을 계산하면, 한 번 나누어서 학습하는 것에 비해 일반화된 성능을 얻을 수 있음
- Adaboost 모델을 학습데이터로 학습한 후 KFold 교차검증을 통해 검증을 진행함
- Fold 중 정확도가 가장 높은 Fold를 최적 모델로 선정함
- 선정된 최적의 모델을 활용하여 평가 데이터에 대해 평가를 진행하며, 그 결과인 혼동행렬은 아래와 같음
- 혼동행렬을 통해 확인할 수 있듯이, 총 3,453건의 데이터 중 3,450건에 대해서 정교하게 분류하였으며, 99.91%의 정확도를 확인할 수 있음



[그림 12] Adaboost 모델 평가 결과

□ AI 모델링 결과

- Decision Tree, Random Forest 및 Adaboost 3개의 모델 중 가장 성능이 좋은 모델은 Random Forest임을 확인할 수 있음

	precision	recall	f1-score	accuracy
DecisionTree	0.956522	1.000000	0.977778	0.999710
RandomForest	1.000000	1.000000	1.000000	1.000000
AdaBoost	1.000000	0.863636	0.926829	0.999131

[그림 13] 모델별 훈련결과

□ 문제점 해결 여부

- 사출품 품질분류에 AI 분석 방법론을 적용한 결과 알고리즘별로 각 데이터셋에 대해 비슷한 성능을 보였음
- Decision Tree, Random Forest 및 Adaboost 중 Random Forest 모델이 여러 평가 지표에서 근소하게 우수한 성능을 보임을 확인하였음
- 앞서 설명된 바와 같이 모델의 성능은 정확도도 중요하지만, 실제 공정에서 발생하는 불량품을 얼마나 잘 예측하느냐를 평가하는 재현율(recall) 및 발생하는 양품을 얼마나 잘 예측하느냐를 평가하는 정밀도 (precision)도 매우 중요하다고 볼 수 있음
- 따라서, 본 가이드북에서는 재현율과 정밀도의 조화평균인 F1-score도 별도로 계산하였으며, F1-score 역시

Random Forest 모델이 가장 성능이 좋았음

- 이러한 점을 고려할 때, 사용된 AI 방법론 중 Random Forest 모델이 사출공정 품질 분류공정에서 가장 훌륭한 성능을 가진다고 할 수 있음

2. 타현장 적용 방안

가. 유사 타현장의 「AI 데이터셋」 분석 적용 방안

□ 본 분석이 적용 가능한 제조 현장 소개

- 본 분석에서는 사출 공정 내 온도, 압력, 속도와 관련된 여러 데이터를 활용하여 사출공정 품질 분류 모델 구축을 진행하였음
- 여러 의사결정나무 계열 알고리즘들을 사용하여 모델을 개발하고, 최적의 모델을 선정하였음. 이러한 분석 프로세스를 통해 사용자는 데이터 기반의 의사결정을 내릴 수 있게함
- 본 분석의 결과물인 사출공정 품질 분류 모델 개발 및 구축을 위해선 공정의 주요 요소들에 대한 데이터 수집 및 관리가 중요함. 본 분석에 적용된 모델들은 의사결정나무 계열 모델들이며, 의사결정나무 계열 모델들의 가장 큰 장점으로는 각 변수 별 중요도를 계산할 수 있다는 점과 중요도를 기반으로 변수별 품질 최적구간을 역산할 수 있다는 점이 있음
- 따라서, 의사결정나무 계열 모델들의 장점과 본 분석에 사용된 변수들의 특징을 활용한다면 품질 분류 모델 구축이 필요한 다른 사출업종 기업들이 손쉽게 활용할 수 있을 것으로 기대됨
- 또한, 온도, 압력, 속도 등의 변수들은 사출이 아닌 타 공정에서도 중요함. 이러한 점에서 본 분석과 동일한 프로세스를 활용하면 타 공정에서도 품질분류 모델 개발 및 구축이 가능할 것으로 예상됨

□ 본 분석을 원용하여 타 제조 현장 적용 시, 주요 고려사항

- 먼저 각 공정 변수를 수집할 수 있는 시스템 구축이 선행되어야함. 특히, 본 분석을 효과적으로 진행하기 위해 필요한 공정 변수들로는 제품 및 공정의 기준정보, 제품의 생산 및 품질 정보, 설비의 공정 및 환경 데이터가 있음
- 제품 기준정보는 제품을 식별할 수 있는 상세 정보인 제품번호, 제품명, 제품군 등이 있으며, 공정 기준정보는 공정을 식별할 수 있는 상세 정보인 공정번호, 공정명, 공정순서 등이 있음
- 제품 및 공정마다 예측 결과가 달라질 수 있으므로 정확한 제품 기준정보와 공정 기준정보를 통해 제품 및 공정 별 AI 모델을 구축 할 수 있음
- 생산 및 품질 정보는 공장에서 생산되는 제품의 생산 실적 정보와 품질 정보로 구성됨. 생산 실적 정보로는 각 제품에 대한 생산 실적 정보인 제품 정보, 공정 정보, 생산 일시, 생산 수량, 작업 번호 등이 있으며, 품질 정보로는 각 생산품에 대한 품질 정보인 합격 여부, 불량 유형 등이 있다. 제품의 생산 실적 정보와 품질 정보가 누락되지 않도록 주의해야함
- 환경 데이터로는 제품 생산에 영향을 주는 설비 외적인 데이터인 주변 온도, 습도 등이 있음. 환경 데이터가 달라지면 AI 모델의 정확한 예측이 어려울 수 있음. 환경 데이터의 변화에 따른 수집과 환경 데이터를 변하지 않도록 고정하는 등의 해결방안을 마련하여 데이터를 수집해야 함
- 설비 공정 데이터는 제품을 생산하면서 발생하는 설비의 공정 데이터이며, 제품 정보, 온도, 압력, 속도 등의 설비 공정 데이터가 있음. 설비 공정 데이터와 품질 정보가 잘못 매칭되어 있다면 부정확한 예측 결과를 초래할 수 있기 때문에 설비 공정 데이터와 품질 정보간 정확한 매칭을 통한 모델 구축으로 정확한 품질 예측을 할 수 있음
- 분석과정에서 데이터의 결측 비율로 인해 또는 유의미하지 않다고 판단되어 제거한 변수가 실제 공정에서 중요한 변수일 수도 있기 때문에 현장 작업자의 의견을 수렴하여 제거 여부를 결정 해야함