



# 프로젝트 4\_3조\_분석기획서\_rev1



팀명: yo yo yo yo yo

팀원: 정은서, 이예솔, 김연진, 장일준, 이다경

## 1 - 비즈니스 배경 및 문제 식별

- 각 불량에 대한 발생 원인과 대책이 정의되어 있으나, 작업자의 경험에 의해 설비를 운영하고 있어 이를 적용하여 해결하지 못하고 있는 상황
- 체계적인 관리 하에 일정한 공정 환경 및 공정 변수를 관리해 불량에 대응하는 것이 필요
  - 공정 개선 필요
  - 공정 개선을 위한 데이터 분석

## 2 - 주제 확정 및 해결방안 수립

### 2.1 - 후보 분석주제 확정

No.	분석주제명	비즈니스 문제	활용방안	기대효과
1	불량/정상 분류 모델 활용 데이터분석	금형 작동 시 불량 발생으로 리스크 있음.	불량을 야기하는 변수 파악 후 공정에 활용.	리스크 감소. 공장 효율 증가. 수익 증가.
2	양산공정에서 불량/정상 분류 모델 활용 데이터분석	고속 양산시 불량 발생으로 리스크 있음.	불량을 야기하는 변수 파악 후 공정에 활용.	리스크 감소. 공장 효율 증가. 수익 증가.
3	생산 사이클타임 감소 데이터 분석	제품 한 개의 생산 시간이 길면 공장 효율 떨어짐.	생산시간을 길게하는 변수 파악 후 공정에 활용.	생산 시간 감소. 공장 효율 증가.

No.	분석주제명	비즈니스 문제	활용방안	기대효과
				수익 증가.

→ [2번] 양산공정에서 불량/정상 분류 모델 활용 데이터분석 선정

## 2.2 - 해결방안 수립

- 양산상태인 데이터만 분석에 사용
  - 사탕주조(try shot): 금형의 온도를 높이기 위한 저속의 주조 공정
  - 사탕신호가 있는 데이터는 100% 불량이다.
- 사탕신호 Nan값인 데이터만 추출
- 분류 모델 적용
- 불량에 영향 주는 변수 확인
- 유지해야되는 변수 값 현장 담당자에게 전달
- 공정 개선

## 3 - 데이터 정의 및 식별

- data\_week4.csv
- 92015 행, 32 열
- 데이터 수집 날짜: 2019-01-02 ~ 2019-03-31
- 라인명: 전자교반 3라인 2호기
- 제품명: TM Carrier RH
- 금형명: TM Carrier RH-Semi-Solid DIE-06

### 3.1 - 컬럼 설명 표

#	컬럼명	컬럼 설명	데이터 타입(수정)	데이터 값 범위	결측값 갯수
0	Unnamed: 0	인덱스 표시	정수	0 ~ 92014	
1	line	작업라인	문자	전자교반 3라인 2호기	
2	name	제품명	문자	TM Carrier RH	
3	mold_name	금형명	문자	TM Carrier RH-Semi-Solid DIE-06	
4	time	수집시간	문자	2019-01-02 ~ 2019-03-31	
5	date	수집일시	문자	00:00:00 ~ 23:59:59	
6	count	일자별 제품 생산 번호	정수	1 ~ 334	

#	컬럼명	컬럼 설명	데이터 타입(수정)	데이터 값 범위	결측값 갯수
7	working	가동여부	문자	가동, 정지	1
8	emergency_stop	비상정지	문자	ON	1
9	molten_temp	용탕온도	실수	0.0 ~ 735.0	2261
10	facility_operation_cycleTime	설비 작동 사이클 시간	정수	0 ~ 457	
11	production_cycletime	제품 생산 사이클 시간	정수	0 ~ 485	
12	low_section_speed	저속구간속도	실수	0.0 ~ 65535.0	1
13	high_section_speed	고속구간속도	실수	0.0 ~ 388.0	1
14	molten_volume	용탕량	실수	0.0 ~ 2767.0	45130
15	cast_pressure	주조압력	실수	41.0 ~ 348.0	1
16	biscuit_thickness	비스킷 두께	실수	0.0 ~ 422.0	1
17	upper_mold_temp1	상금형온도1	실수	18.0 ~ 1449.0	1
18	upper_mold_temp2	상금형온도2	실수	15.0 ~ 4232.0	1
19	upper_mold_temp3	상금형온도3	실수	42.0 ~ 1449.0	313
20	lower_mold_temp1	하금형온도1	실수	18.0 ~ 369.0	1
21	lower_mold_temp2	하금형온도2	실수	20.0 ~ 1449.0	1
22	lower_mold_temp3	하금형온도3	실수	299.0 ~ 65503.0	313
23	sleeve_temperature	슬리브온도	실수	23.0 ~ 1449.0	1
24	physical_strength	형체력	실수	0.0 ~ 65535.0	1
25	Coolant_temperature	냉각수 온도	실수	16.0 ~ 1449.0	1
26	EMS_operation_time	전자교반 가동시간	정수	0 ~ 25	
27	registration_time	등록일시	문자 → 시간	2019-01-02 16:45:06 ~ 2019-03-31 23:59:48	
28	passorfail	불량판정(0: 양품 1:불량)	실수 → 불리언	0, 1	1
29	trysot_signal	사탕신호	문자	D	90096
30	mold_code	금형코드 (7개 금형에 관한 코드)	정수 → 문자	8412, 8413, 8573, 8576, 8600, 8722, 8917	0
31	heating_furnace	가열로	문자	A, B	49146

### 3.2 - 추가 컬럼 설명 표

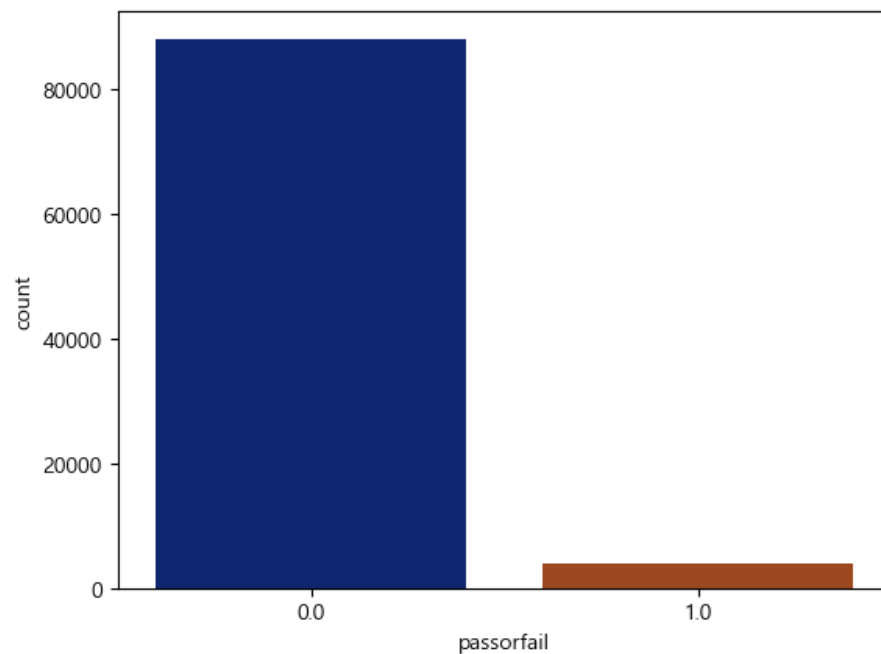
#	컬럼명	컬럼 설명	데이터 타입(수정)	데이터 값 범위
1	month	월	정수	1~3

#	컬럼명	컬럼 설명	데이터 타입(수정)	데이터 값 범위
2	day	일	정수	1~31
3	hour	시	정수	0 ~ 23
4	minute	분	정수	0 ~ 59
5	second	초	정수	0 ~ 59

## 4 - 분석 정의

### 4.1 - 목표변수

- passorfail  
불량판정 (0: 양품 1:불량) → 양품 87998, 불량 4017
- 불균형 데이터임.  
약 95.6 : 4.4 비율 → 샘플링 기법 사용
- nan 값이 1개 존재하여 제거할 예정.



### 4.2 - 샘플링 기법

- SMOTE
- 랜덤오버샘플링
- 모델 자체 가중치 사용

### 4.3 - 사용 데이터

- tryshot\_signal == nan  
사탕코드가 비어있는 데이터 사용  
원본 데이터에서 사탕코드가 'D'인 1919개의 행을 제외한다.
- passorfail 비중 (정상: 87998, 불량: 4017 → 2098, 1919행 제거)

## 4.4 - 사용 모델

- 분류 모델 사용
  - Cat Boost
  - LightGBM
  - Random Forest
  - Logistic Regression

## 4.5 - 평가기준 정의

- 불량품 리스크 크다.  
Recall값에 비중을 둔다.
- ROC AUC
- G-mean

## 5 - 분석 결과 활용 방안

- 금형 공정 관리자와 미팅을 통해 분석 결과 전달
- 주요 수치들 확인 후 공정 개선
- 불량 발생 감소
- 공장 효율 증가
- 수주처 만족, 수입 증가