



분석 기획서

팀명: 2.5조

팀원: 정한렬, 안상후, 오서연, 윤대웅, 이승학, 정아영

1. 비즈니스 배경 및 문제 식별

1.1 고객 분석 요청 배경

- 불량 발생으로 인한 비용 증가
 - 주조 공정에서 불량률이 높아지면 재료 낭비 및 재작업 비용이 발생
 - 불량률을 낮춰 비용 절감과 생산성 향상을 목표
- 공정 제어의 비체계성
 - 현재 공정이 작업자의 경험에 의존하여 운영되고 있으며, 이를 데이터 기반으로 체계화하지 못하고 있음
 - 데이터 기반 관리 시스템을 도입해 설비 가동률과 품질을 개선
- 공정 변수 최적화 필요성
 - 주조 공정에서는 용탕온도, 가열로 온도, 금형 온도와 같은 변수들이 품질에 큰 영향

- 각 변수와 불량 간의 인과관계를 분석하여 공정을 최적화하고, 자동화된 제어를 구축하려는 니즈

1.2 고객 KPI

- 품질 개선, 비용 절감, 생산성 향상을 목표로 한다.

2. 해결 방안 수립 및 주제 확정

- **인과 관계 분석:** 설명 변수의 변화가 먼저 일어나고 불량이 발생했는지, 혹은 불량이 발생하고 설명 변수의 변화가 일어났는지를 확인하여 변수들 간의 선행 관계를 분석한다.
- **분류 분석**
 - 불량 판별 모델을 생성하여 변수들 간의 선행 관계 고려해 shift 파생 변수를 추가한다.
 - 불량이 다수 발생하는 시간대의 데이터만 가지고 모델 생성하는 방안도 고려한다.
- **공정 최적화:** 불량을 야기하는 주요 변수들의 조건을 파악해 불량률을 줄이는 방향성을 제시한다.

3. 데이터 정의 및 식별

92015 rows × 31 columns

변수명	결측치 수	데이터 타입	Unique 값
작업라인	0	object	1
제품명	0	object	1
금형명	0	object	1
수집날짜	0	object	85
수집시각	0	object	54640
일자별제품생산번호	0	int64	334
가동여부	1	object	2
비상정지	1	object	1

용탕온도	2261	float64	113
설비작동사이클시간	0	int64	195
제품 생산 사이클시간	0	int64	202
저속구간속도	1	float64	123
고속구간속도	1	float64	230
용탕량	45130	float64	121
주조압력	1	float64	124
비스킷 두께	1	float64	67
상금형온도1	1	float64	317
상금형온도2	1	float64	232
상금형온도3	313	float64	117
하금형온도1	1	float64	345
하금형온도2	1	float64	323
하금형온도3	313	float64	314
슬리브온도	1	float64	590
형체력	1	float64	106
냉각수 온도	1	float64	35
전자교환 가동시간	0	int64	5
등록일시	0	object	91722
불량판정	1	float64	2
사탕신호	90096	object	1
금형코드	0	int64	7
가열로	49146	object	2

- **생성 규칙** : 분, 초 단위로 불규칙하게 생성
- **관리 상태** : 2019-01-02부터 2019-03-31까지 수집된 데이터 (총 85일)

4. 분석 정의

4.1 행 드롭 및 파생변수 추가

1. 의미없는 컬럼 드롭

- Unnamed:0, 일자별제품생산번호(Count)

2. 유효값이 하나인 컬럼 드롭

- 작업라인, 제품명, 금형명, 비상정지

3. 시간 변수 통합

- 수집날짜, 수집시각을 드롭 후, '등록일시'를 date_time으로 바꿔 필요시 변환해 사용

4. 사탕신호 드롭

- 사탕신호는 테스트 주조물로 불량 여부를 판별하는 것이므로 예측 모델에 사용하는 것이 적합하지 않음

3. 온도별 파생 변수 추가

- 오전 N시간 이동평균 임계값 미만 : True / False
- 오후 N시간 이동평균 임계값 미만: True / False
- N시간 이동 평균

4. 독립 변수 threshold의 이상 이하 여부

4.2 분석기간 정의

Train set	Test set
2019-01-02~2019-03-30	2019-03-31
2019-01-02~2019-03-29	2019-03-30
2019-01-02~2019-03-28	2019-03-29
2019-01-02~2019-03-27	2019-03-28
2019-01-02~2019-03-26	2019-03-27
2019-01-02~2019-03-25	2019-03-26
2019-01-02~2019-03-24	2019-03-25

2019-01-02~2019-03-23	2019-03-24
2019-01-02~2019-03-22	2019-03-23
2019-01-02~2019-03-21	2019-03-22
2019-01-02~2019-03-20	2019-03-21
2019-01-02~2019-03-19	2019-03-20
2019-01-02~2019-03-18	2019-03-19
2019-01-02~2019-03-17	2019-03-18

4.3 분석모델

머신러닝	시계열 모델
XGBoost	Prophet
LightGBM	LSTM
CatBoost	
RandomForest	

추후 변경될 수 있음

4.4 평가 기준 정의

- **Precision, Recall:** 2종 오류를 줄이기 위해 recall을 우선적으로 고려하여 사용한다.
- **ROC - AUC:** TPR과 FPR을 고려하여 모델 성능을 비교하기 위해 사용한다.
- **G-Mean:** 클래스 불균형이 심하기 때문에 G-Mean을 통해 민감도와 특이도의 균형을 평가한다.

5. 분석 결과 활용 방법

- **설명 변수의 인과 관계 분석**을 통해 공정 관리의 초점을 어디에 두어야 할지 결정할 수 있음. 또한, 공정 관리의 타임라인을 최적화 할 수 있다.
- **불량 판별 모델**을 통해 불량 예측의 정확도를 높이고, 불량이 다수 발생하는 시간대에 집중하여 **공정을 최적화**하여 불량률을 줄일 수 있음. 이를 통해 **비용 절감**이 가능하다.
- **공정 최적화 결과**를 통해 공정의 목표값 설정을 화용해 실시간 모니터링과 공정 관리를 강화할 수 있음. 또한, 공정 매뉴얼을 개선하거나 공정 자동화 시스템의 제어값 설정 가능. 뿐만 아니라, **품질을 개선**할 수 있다.