

제조 공정 불량 분석 및 예측 모델 구축

개인프로젝트

1. 프로젝트 개요

본 프로젝트는 Bosch 생산 라인 데이터를 기반으로 제품 생산 과정에서 발생하는 불량 패턴을 분석하고, 불량이 집중되는 공정 구간을 도출하여 개선 가능성을 탐색하는 것을 목표로 합니다.

단순 불량 예측을 넘어, 공정 흐름 관점에서 불량 발생 구조를 이해하고 개선 가능성을 도출하는 데 초점을 두었습니다.

2. 문제 정의

대규모 제조 공정에서는 다음과 같은 문제가 발생합니다.

- 불량 제품이 어떤 공정 단계에서 발생하는지 파악하기 어려움
- 공정 데이터는 방대하지만 문제 구간을 특정하기 어려움
- 불량 예방을 위한 관리 기준 수립이 어려움

따라서 본 프로젝트에서는 다음 질문을 핵심 문제로 설정했습니다.

-> 불량 제품은 정상 제품과 다른 공정 흐름을 가지며, 특정 문제 공정을 찾을 수 있는가?

3. 데이터 설명

Kaggle (Bosch Production Line Performance) 데이터를 활용하였으며,
각 제품이 공정을 통과하며 기록된 측정 데이터와 최종 불량 여부가 포함되어 있습니다.

주요 데이터 정보:

- 공정 단계별 측정 시간 정보
- 공정 통과 여부 및 측정 기록
- 최종 제품 불량 여부(Response)

대규모 공정 로그 데이터를 기반으로 분석을 진행했습니다.

4. 분석 접근 방식

문제를 해결하기 위해 다음 단계로 분석을 진행했습니다.

1. 불량 데이터 분포 및 공정 시간 패턴 분석
2. 공정 통과 횟수 및 측정 누락 패턴 비교
3. 공정 station별 불량률 분석
4. 불량 집중 공정 라인 탐색
5. 문제 공정 이전 공정 흐름 분석
6. 분석 결과 기반 예측 모델 구축

5. 핵심 분석 결과

분석 결과 다음과 같은 패턴을 확인했습니다.

- 불량 제품은 정상 제품 대비 공정 흐름이 불안정한 경향 존재

- 공정 시간이 극단적으로 짧거나 긴 제품에서 불량률 증가
- 특정 공정 라인 구간에서 불량이 집중됨

특히, L3 라인의 S32 공정에서 평균 대비 높은 불량률 발생을 확인했으며,

이는 단일 공정 문제가 아니라 **공정 흐름 구간 전체 영향 가능성**을 시사합니다.

6. 문제 해결 과정

문제 1. 방대한 공정 데이터에서 문제 구간을 특정하기 어려운 문제

문제 상황

공정 데이터에 수천 개의 측정 정보가 포함되어 있어
문제가 발생하는 공정 단계를 직접적으로 파악하기 어려웠습니다.

해결 방법

- 공정 흐름을 요약할 수 있도록 공정 시간, 공정 통과 횟수, 측정 누락 비율 등
의 지표 생성
- station 단위 불량률 분석 수행
- 공정 라인 단위 비교를 통해 문제 구간 탐색

해결 결과

특정 공정 라인 구간에서 불량이 집중됨을 확인하고,
L3_S32 공정을 주요 문제 구간 후보로 도출했습니다.

문제 2. 단일 공정 문제인지 공정 흐름 문제인지 판단하기 어려운 문제

문제 상황

문제 공정이 발견되었으나 해당 공정 자체의 문제인지, 전 공정에서 누적된 영향인지 구분하기 어려웠습니다.

해결 방법

- 문제 공정을 거친 제품들의 이전 공정 흐름 분석
- 동일 공정 흐름을 거친 제품들의 불량률 비교
- 공정 구간 단위 문제 발생 가능성 검토

해결 결과

불량이 특정 공정 하나가 아니라
공정 흐름 구간 전체의 영향으로 발생할 가능성이 높음을 확인했습니다.

문제 3. 분석 결과를 예측 모델에 반영하기 어려운 문제

문제 상황

공정 분석 결과를 예측 모델에 어떻게 활용할지 명확하지 않았습니다.

해결 방법

- 공정 소요 시간, 공정 통과 패턴, 문제 공정 방문 여부를 feature로 구성
- 분석에서 도출한 문제 공정 정보를 모델 입력 변수로 추가
- 모델 학습 후 feature 중요도 분석 수행

해결 결과

공정 시간 및 문제 공정 방문 여부가 실제 불량 예측에 중요한 요소로 활용됨을 확인했습니다.

7. 개선 방향 및 프로젝트 의의

분석 결과를 바탕으로 다음과 같은 개선 방향을 제안할 수 있습니다.

- 문제 공정 구간 설비 및 조건 점검 필요
- 해당 공정 진입 전 품질 검사 강화 가능
- 공정 흐름 이상 감지 기반 품질 모니터링 적용 가능

본 프로젝트는 데이터 분석을 통해

공정 흐름 기반 불량 발생 구조를 파악하고,

문제 구간 도출 및 예측 모델 구축까지 연결했다는 점에서 의의를 가집니다.

8. 사용 기술 및 개발 환경

사용 기술

- Python
- Pandas, NumPy
- Matplotlib, Seaborn
- LightGBM
- Scikit-learn

개발 환경

- Google Colab 기반 데이터 분석 환경
- Jupyter Notebook 기반 분석 및 모델링 수행
- GitHub를 통한 코드 관리 및 프로젝트 기록 관리

GitHub Repository

https://github.com/rjaekawpxm1-netizen/portfolio/tree/main/Projects/bosch_production