

전해탈지 공정 데이터 기반 머신러닝 제조분야 적용 연구 – 데이터 분석 및 시각화

정유현, 이미진, 곽민서

Machine Learning Application in Manufacturing Based on Electrolytic Degreasing Process Data – Data Analysis and Visualization

Jeong Yuhyeon,

요 약

본 연구에서는 전해탈지 공정에서 수집된 센서 데이터를 활용하여 품질 예측 모델 구축을 위한 데이터 분석 및 시각화를 수행하였다. KAMP(Korea AI Manufacturing Platform)에서 제공하는 전해탈지 공정 데이터셋을 대상으로 pH, 온도(Temp), 전류(Current) 등 주요 공정 변수의 분포 특성과 상관관계를 분석하였으며, 데이터 품질 지수(완전성, 유일성, 유효성)를 평가하였다. 분석 결과, 데이터는 높은 품질 수준(완전성 100%, 유효성 100%, 유일성 99.64%)을 보였으며, 공정 변수 간 상관관계는 미약하나 독립적인 품질 영향 인자로 작용함을 확인하였다. 본 연구는 중소기업 제조 현장에서 AI 기반 품질 예측 시스템 구축을 위한 데이터 기반 접근법의 실증적 근거를 제공한다.

Abstract

This study performs data analysis and visualization to build a quality prediction model using sensor data collected from electrolytic degreasing processes. Using the electrolytic degreasing process dataset provided by KAMP (Korea AI Manufacturing Platform), we analyzed the distribution characteristics and correlations of key process variables such as pH, temperature (Temp), and current (Current), and evaluated data quality indices (completeness, uniqueness, validity). The analysis results showed high data quality levels (completeness 100%, validity 100%, uniqueness 99.64%), and confirmed that the correlations between process variables were weak but acted as independent quality influencing factors. This research provides empirical evidence for a data-driven approach to building AI-based quality prediction systems in small and medium-sized manufacturing sites.

Key words

electrolytic degreasing, manufacturing data analysis, data quality assessment, exploratory data analysis, process visualization, machine learning for manufacturing

*소속, Email, **소속, Email(교신저자표시), ...

※ 지원기관표기(사사표기)

I. 서 론

표면처리 공정 중 하나인 전해탈지는 금속 표면의 오염물 제거와 표면 활성화를 위한 필수 전처리 공정이다[1]. 전해탈지 공정에서는 pH, 온도, 전류밀도, 공정시간 등 다양한 변수들이 복합적으로 작용하여 최종 제품의 품질에 영향을 미친다. 특히 온도가 설정치를 벗어나면 피도금물 표면 손상이 발생할 수 있으며, 전류밀도와 공정 시간의 부적절한 설정은 수소취성 등의 품질 문제를 야기할 수 있다[2].

중소기업 제조 현장에서는 다양한 공정 데이터가 수집되고 있으나, 이를 체계적으로 분석하고 활용하는 사례는 제한적이다. 데이터 분석 역량 부족과 AI 활용 경험 부재가 주요 원인으로 지적된다[3]. 이에 본 연구에서는 전해탈지 공정 데이터를 대상으로 체계적인 탐색적 데이터 분석(EDA)을 수행하고, 데이터 품질 평가 프레임워크를 적용하여 AI 모델 구축을 위한 데이터 준비 과정을 제시한다.

본 연구의 목적은 (1) 전해탈지 공정 데이터의 통계적 특성 파악, (2) 공정 변수 간 상관관계 분석, (3) 데이터 품질 지수 평가, (4) 시각화를 통한 데이터 이해도 향상이다. 이를 통해 중소기업 제조 현장에서 AI 기반 품질 예측 시스템 구축 시 데이터 준비 단계의 모범 사례를 제시하고자 한다.

II. 데이터셋 및 분석 방법

2.1 데이터셋 개요

본 연구에서는 KAMP(Korea AI Manufacturing Platform)에서 제공하는 전해탈지 공정 데이터셋을 활용하였다. 데이터는 2021년 9월 6일부터 10월 27일까지 5초 주기로 수집되었으며, 총 50,094개의 관측값으로 구성된다. 주요 변수는 표 1과 같다.

2.2 분석 환경 및 도구

분석 환경은 Python 3.8.8 기반으로 구축하였으며, 주요 라이브러리는 다음과 같다.

데이터 처리: Pandas 1.19.5, NumPy 1.17.3

시각화: Matplotlib 3.4.2, Seaborn

통계 분석: SciPy

2.3 데이터 품질 평가 방법

ISO/IEC 25012 데이터 품질 표준을 기반으로 다음의 품질 지수를 평가하였다[4]:

(1) 완전성 품질 지수

$$\text{완전성} = \left(1 - \frac{\text{결측치 수}}{\text{전체 데이터 수}}\right) \times 100$$

(2) 유일성 품질 지수

$$\text{유일성} = \frac{\text{유일한 데이터 수}}{\text{전체 데이터 수}} \times 100$$

(3) 유효성 품질 지수

$$\text{유효성} = \frac{\text{유효 범위 내 데이터 수}}{\text{전체 데이터 수}} \times 100$$

유효 범위는 데이터셋 메타데이터에 정의된 공정 조건(pH: 8~12, Temp: 40~52°C, Current: 5~9A)을 기준으로 설정하였다.

표 1. 실험 데이터 세트

Table 1. Experimental data sets

변수명	설명	데이터 타입	측정 범위
Lot	제품 추적 번호	int	1~22
Time	측정 시각	timestamp	H:MM:SS
pH	공정 pH 측정값	float	9.51~11.99
Temp	공정 온도(°C)	float	42.01~51.98
Current	공정 전류(A)	float	5.51~33.0

III. 분석 결과

3.1 기술통계 분석

pH는 평균 10.27로 알칼리성 범위 내에서 관리되고 있으며, 온도는 평균 45.02°C로 설정 범위 중앙

에 분포한다. 전류는 평균 7.97A이나 최대값 33.0A의 이상치가 관찰되었다.

3.2 데이터 분포 분석

그림 1은 주요 변수의 히스토그램을 나타낸다.

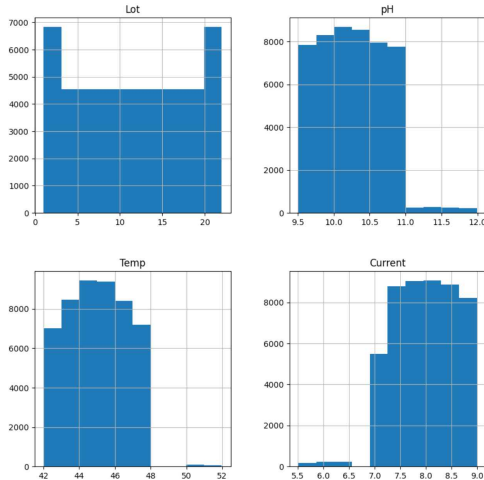


그림 1. 주요 변수의 히스토그램
Fig. 1. Histograms of Key Variables

pH와 온도는 정규분포에 가까운 형태를 보이나, 전류는 약간 우측으로 치우친 분포를 나타낸다. Lot 변수의 경우 1번과 22번에서 높은 빈도를 보이며, 이는 특정 제품군에 대한 집중적인 생산이 이루어졌음을 시사한다.

3.3 상관관계 분석

그림 2는 공정 변수 간 상관관계 분석 결과를 보여준다.

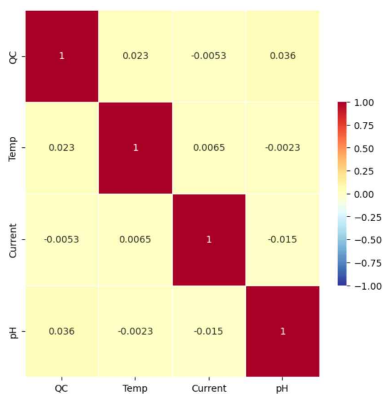


그림 2. 상관관계 히트맵
Fig. 2. Correlation Heatmap

분석 결과, 모든 공정 변수 간 상관관계수의 절대값이 0.03 미만으로 매우 약한 상관관계를 나타냈다. 이는 각 변수가 독립적으로 품질에 영향을 미치는 인자임을 시사한다.

3.4 데이터 품질 평가

표 2는 데이터 품질 지수 평가 결과를 요약한다. 완전성과 유효성은 100%로 매우 우수한 수준이나, 유일성이 99.64%로 일부 중복 레코드가 존재함을 확인하였다. 이는 공정의 정상 상태 구간에서 동일한 조건이 반복 측정되었기 때문으로 판단된다.

표 2. 데이터 품질 지수 평가 결과

Table 2. Data Quality Index Assessment Results

품질 지수	평가 결과	설명
완전성	100%	결측치 0개
유일성	99.64%	pH-Temp-Current 조합 기준
유효성	100%	모든 값이 유효 범위 내
무결성	66.66%	유일성 미만족으로 인한 감점

IV. 결론 및 향후 연구

본 연구에서는 전해탈지 공정 데이터에 대한 체계적인 탐색적 데이터 분석을 수행하였다. 주요 연구 결과는 다음과 같다:

(1) 공정 변수들은 설정 범위 내에서 안정적으로 관리되고 있으며, 일부 이상치를 제외하면 정규분포에 가까운 양호한 분포 특성을 보인다.

(2) pH, 온도, 전류 간 상관관계가 매우 약하여($|r| < 0.03$), 각 변수가 독립적인 품질 영향 인자로 작용함을 확인하였다.

(3) 데이터 품질 평가 결과 완전성 100%, 유효성 100%, 유일성 99.64%로 AI 모델 학습에 적합한 고품질 데이터임을 검증하였다.

본 연구는 중소기업 제조 현장에서 수집된 데이터를 AI 모델에 활용하기 위한 데이터 준비 과정의

모범 사례를 제시하였다는 점에서 의의가 있다.

참 고 문 헌

- [1] 중소벤처기업부, Korea AI Manufacturing Platform(KAMP), "품질 이상탐지, 진단(전해탈지) AI 데이터셋", KAIST, 2021.
- [2] 한국표면공학회, "표면처리 기술핸드북", 2020.
- [3] 김철수, 이영희, "중소제조기업의 AI 도입 장애요인 분석", 한국정보기술학회논문지, 제19권 제2호, pp. 45-56, 2021.
- [4] ISO/IEC 25012:2008, "Software engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Data quality model", 2008.

한글제목	휴먼명조, 17, 장평:90, 자간: -7
저자명	돋움, 11, 장평:90, 자간: 5
영문제목	견명조, 15, 장평:90, 자간: -7
영문저자명	휴먼명조, 10, 장평:90, 자간: 5
요약본문	휴먼명조, 9.2, 장평:90, 자간:-6
영문요약문	영문:Times New Roman, 9.2, 장평:90, 자간:-6
각장제목	휴먼고딕, 11, 장평:90, 자간:-6
본문내용	휴먼명조, 10, 장평:90, 자간:-6
소제목	중고딕 11, 장평:90, 자간:-6
그림캡션	중고딕, 9, 장평:90, 자간:-6
표캡션	중고딕, 9, 장평:90, 자간:-6
식	크기 9, 원정렬, 식번호는 오른 정렬
참고문헌	영문:Times New Roman, 10, 장평:90, 자간:-6

[