

# A Study on Reliability Measurement Method for Data Quality Management Using Digital Transformation-Based Smart Factory Process Data

Heung-Sup Sim\*

\*Professor, Dept. of Digital Twin Software, Dongyang University, Yeongju City, Korea

## [Abstract]

Digital transformation (DX) is driving the implementation of smart factories in manufacturing, contributing to real-time analysis and optimization of process data. However, low data reliability can lead to quality degradation and increased costs, and problems such as missing values, outliers, and duplicate data can arise. In this study, we developed an AI-based data cleansing and quality management framework, applied it to manufacturing process data, and aimed to improve data reliability through anomaly detection using machine learning. Through the analysis of Oil Gasket manufacturing data, we empirically evaluated the reduction of defect rates, improvement of production speed, and optimization of resources. We assessed the core elements of data quality, including accuracy, completeness, consistency, reliability, timeliness, and validity, and enhanced data integrity using AI models such as Random Forest, Autoencoder, KNN, and Multivariate Regression.

As a result of the research, it was confirmed that AI-based data quality management is effective in improving the productivity and quality of manufacturing, and the Data Quality Index (DQI) also improved by 6.5%. Through the use of various AI models, it was confirmed that the Random Forest model, in particular, has excellent performance in classifying defective products. Future research will propose a smart factory operating model through real-time data processing and automation, and present more effective quality management methods by building an AI-based quality management system.

▶ **Key words:** Digital Transformation, Data Quality, Smart Manufacturing, AI-based Cleansing, Anomaly Detection

## [요 약]

디지털 전환(DX)은 제조업에서 스마트팩토리를 구현하여 공정 데이터를 실시간으로 분석하고 최적화하는 데 기여하고 있다. 그러나 데이터 신뢰성이 낮으면 품질 저하와 비용 증가 초래한다.

본 연구에서는 AI 기반 데이터 정제 및 품질 관리 프레임워크를 개발하여 제조 공정 데이터에 적용하고, 머신러닝을 활용한 이상 탐지를 통해 데이터 신뢰성을 개선하고자 하였다. Oil Gasket 제조 데이터 분석을 통해 불량률 감소, 생산 속도 향상, 자원 최적화를 실증적으로 평가하였다. 데이터품질의 핵심 요소인 정확성, 완전성, 일관성, 신뢰성, 적시성, 유효성을 평가하고, 랜덤 포레스트, 오토인코더, KNN, 다변량 회귀 등의 AI 모델을 활용하여 데이터 정합성을 높였다.

연구 결과, AI 기반 데이터품질 관리가 제조업의 생산성과 품질을 향상하는데 주요 인자로 확인하였으며, 데이터품질 지수(DQI) 또한 6.5% 향상되었다. 특히 랜덤 포레스트 모델이 불량품 분류 성능 확인, 향후 연구에서는 실시간 데이터 처리를 통한 스마트팩토리 운영 모델을 제안, AI 기반 품질 관리 시스템을 구축하여 더욱 효과적인 품질 관리 방안을 제시할 예정이다.

▶ **주제어:** 디지털 전환, 데이터품질, 스마트 제조, AI 기반 정제, 이상 탐지

- First Author: Heung-Sup Sim, Corresponding Author: Heung-Sup Sim
- Heung-Sup Sim (mylee911@naver.com), Dept. of Digital Twin Software, Dongyang University
- Received: 2025. 02. 26, Revised: 2025. 03. 20, Accepted: 2025. 03. 31.

## I. Introduction

디지털 전환(DX)은 4차 산업혁명의 핵심 개념으로 자리 잡으며, 제조업의 스마트팩토리 구현을 촉진하고 있다. 스마트팩토리는 공정 데이터의 실시간 수집과 분석을 통해 제조 공정을 최적화하고 생산성을 향상시키는 것을 목표로 한다. 센서 및 IoT 기기를 활용하여 수집된 데이터는 제품 품질 관리, 공정 효율성 증대, 유지보수 최적화 등의 중요한 역할을 한다[1].

그러나 스마트팩토리 환경에서 생성되는 대량의 공정 데이터는 높은 신뢰성을 요구한다. 데이터의 신뢰성이 낮으면 잘못된 의사 결정으로 이어질 수 있으며, 이는 제품 결함 증가, 생산 비용 상승 및 품질 저하를 초래할 수 있다. 데이터품질 문제는 주로 결측 데이터, 이상값, 중복 데이터 및 노이즈 발생에서 기인한다. 이에 따라 제조업에서는 데이터 정제 기술을 적극적으로 도입하여 데이터 신뢰성을 보장하는 것이 필수적이다[2].

스마트팩토리 공정 분야도 빅데이터 분석 기반으로한 모델성능평가 부분의 인공지능(AI)기술을 활용한 데이터품질 관리 기법이 주목받고 있다. 머신러닝 기반의 이상 탐지 기술은 공정 데이터를 분석후 이상치를 감지하고 자동으로 보정하는 기능을 제공하며, 이는 공정 신뢰성과 데이터 품질을 높이는 데 중요한 역할을 한다. 본 연구에서는 AI 기반의 데이터 정제 및 품질 관리 프레임워크를 개발하고, 이를 제조 공정 데이터에 적용하여 데이터 신뢰성이 제품 품질과 공정 최적화에 미치는 영향을 공정 데이터로 분석하고자 한다[3].

본 연구의 주요 목적은 다음과 같다.

공정 데이터품질에 영향을 미치는 주요 변수를 식별하고 데이터의 패턴을 분석하여, AI 기반 데이터 정제 및 이상 탐지 프레임워크를 개발하여 데이터 품질을 향상을 검증하고, 데이터품질과 제조 공정 효율성 간의 관계를 실증적으로 활용한다.

제안된 모델을 실제 제조 데이터에 적용하여 그 효과를 평가하고, 공정 최적화 방안을 고려한 실시간 데이터품질 평가 및 보정 시스템을 설계하고 성능을 분석하여 현장 적용 가능성을 검토한다.

특히, 본 연구에서는 데이터품질이 제조업의 주요 생산 지표(예: 불량률 감소, 생산 속도 향상, 원자재 활용 최적화 등)에 미치는 영향을 다각도로 분석할 것이다. 이를 통해 AI 기반 데이터 정제 및 품질 관리 기법이 제조업 전반에서 적용될 수 있는 실무적인 대책을 제안하고자 한다. 향후 연구에서는 실시간 데이터 처리 및 공정 자동화 기술

과 연계하여 보다 효과적인 스마트팩토리 운영 모델을 구축하는 방향으로 발전시킬 것이다[4].

연구에 활용한 데이터의 개요를 요약하면 2024년 스마트제조혁신추진단의 “선도형 스마트공장 구축지원 사업”의 (주)대우비엔지 실증 데이터를 기반으로 제조 공정 데이터를 분석하여 데이터 품질을 검증하고 개선 전략을 제안한다. 분석에 사용된 데이터는 제조업에서 Oil Gasket 생산 공정에서 수집된 데이터를 활용하였다.

## II. Preliminaries

### 1. Related works

#### 1.1 Analyzing manufacturing process data

본 연구에서는 자동차 부품 생산 데이터를 분석 대상으로 삼았으며, 주요 변수로 실압력(MPA), 고정 시간(fix\_time), a 속도(a\_speed), 하중 시간(load\_time), b 속도(b\_speed), 고압 시간(highpressure\_time), 탕구 두께(c\_thickness) 등을 고려하였다. 해당 데이터는 약 34,139 개의 샘플로 구성되어 있으며, 탕구 두께가 [21,27] 범위를 벗어나면 불량으로 간주하였다.

Table 1. Summary of Key Findings and Visualization Criteria from the Paper

Item	Paper Findings and Visualization Criteria
Analyzed Data	-Manufacturing Data for Automotive Components (Oil Gasket)
Main Process Variables	-fix_time, a_speed, b_speed, mpa,load_time, highpressure_time
Quality Variable	Gate Thickness (c_thickness)
Anomaly Detection and Key Findings	-Defect rate increases by over 30% when load_time exceeds 20.1 seconds. -Quality variability significantly increases at specific intervals of highpressure_time.
Criteria for Process Variables Impact Analysis	-Each variable categorized into low, middle, and high based on quantiles for analysis.
Boxplot Visualization Criteria	-Quantile-based categorization into three intervals. - Defect thresholds visually presented (thickness $\leq 21$ or $\geq 27$ ).

분석 결과 전체 데이터 중 11%가 불량 제품으로 판명되었다.

탕구 두께(c\_thickness)는 공정 변수들의 변화에 직접적인 영향을 받으며, 특히 a\_speed, b\_speed, 실압력(MPA) 등의 변수와 강한 상관관계를 보였다.

분석 결과 b\_speed 값이 1.732 이상일 경우 제품 두께

가 급격히 감소하는 경향을 보였으며, 실압력이 일정 수준 이하로 떨어질 경우 품질 변동성이 증가하는 패턴을 확인할 수 있었다.

또한, 공정 데이터의 이상값을 탐지하기 위해 각 변수의 분포를 분석하였으며, 특정 구간에서 비정상적인 데이터가 다수 존재함을 확인하였다. 예를 들어, 하중 시간(load\_time)이 20.1 초 이상인 경우 불량률이 30% 이상 증가하는 현상이 나타났으며, 고압 시간의 특정 값에서 품질 변동성이 심화되는 패턴을 보였다.

공정변수의 영향력 분석 - 각 공정변수의 수준에 따른 탱구 두께의 분포공정 변수들이 탱구 두께에 미치는 영향을 파악하기 위해 시각적 분석을 수행하였다. 각 변수를 low, middle, high의 3개 구간으로 나누어 탱구 두께의 변화를 상자그림(Boxplot)으로 나타냈다.

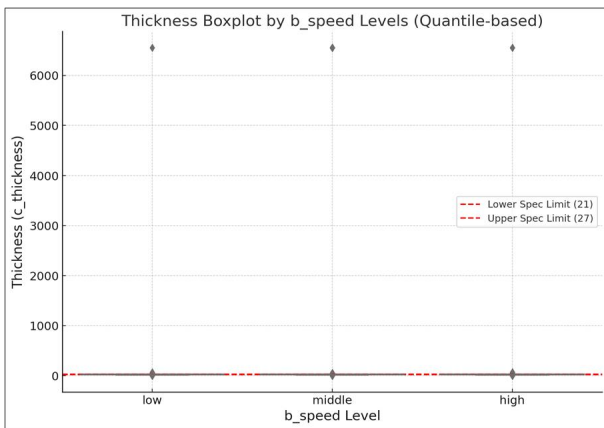


Fig. 1. Thickness Boxplot By B\_speed Levels (Quantile-Based)

분석 결과, b\_speed가 1.7 이하인 경우 제품 두께가 상대적으로 안정적이었으나, 1.7을 초과하면 급격히 감소하는 현상을 보였다. 또한, 실압력(mpa)이 75 이상일 때 제품 두께의 변동성이 커지는 경향을 보였다.

탱구 두께가 [21,27] 범위를 벗어나는 경우 불량률이 증가하는 것으로 나타났으며, 하중 시간(load\_time)이 20.1초를 초과할 경우 불량률이 30% 이상 상승하는 패턴을 확인하였다. 또한, 공정변수들 간의 상관계수를 분석한 결과, 실압력과 a 속도 간의 상관관계가 가장 강하게 나타났다. 상관행렬 분석을 위해 corrplot 패키지를 활용하여 시각화하였으며, corrplot.mixed() 함수를 적용하여 상관계수 수치와 상관계수의 크기를 타원의 모형과 색깔로 표현하였다.

분석 결과, 타원의 모양이 원에 가까울수록 약한 상관관계를 나타내며, 직선 형태에 가까울수록 강한 상관관계를 나타내는 것으로 확인되었다.

(1) 강한 상관관계를 가지는 변수를 고려한 경우.

실압력(mpa)과 a\_speed: 0.86 (강한 양의 상관) 실압력이 증가할수록 a 속도도 함께 증가하는 경향이 나타남.

탱구두께(c\_thickness)와 실압력(mpa): -0.69 (강한 음의 상관) 실압력이 증가할수록 탱구 두께가 얇아지는 경향이 있다. 탱구두께(c\_thickness)와 a\_speed: -0.62 (강한 음의 상관) a 속도가 증가할수록 탱구 두께가 얇아지는 경향이 있다.

(2) 중간 정도의 상관관계를 가지는 변수를 고려한 경우 탱구두께(c\_thickness)와 load\_time: 0.24 (약한 양의 상관) 하중 시간이 길어질수록 탱구 두께도 증가하는 경향이 있으므로, fix\_time과 load\_time: 0.44 (중간 정도의 양의 상관) 고정 시간이 길수록 하중 시간도 증가하는 경향이 있다.

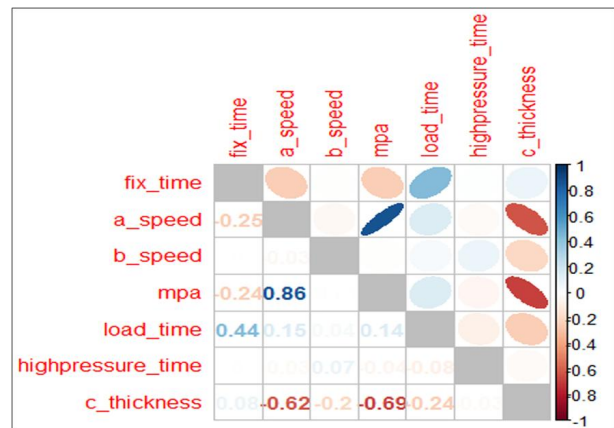


Fig. 2. Observation Item Variable Correlation Results

(3) 상관관계가 약한 변수를 고려한 경우

b\_speed는 대부분의 변수와 상관관계가 미미함.

highpressure\_time도 대부분의 변수와 낮은 상관관계를 가진다. 탱구두께(c\_thickness)와 highpressure\_time은 0.03으로 사실상 관계가 없다.

특히, 탱구 두께와 가장 많은 선형적 연관성을 포함한 변수는 a 속도, 실압력으로 분석되었으며, 두 변수 모두 음의 상관을 보였다. a 속도나 실압력이 커질수록 탱구 두께는 얇아지는 경향성이 있는 것으로 확인되었다.

이러한 연관성을 보다 정확히 분석하기 위해 선형회귀모형을 적용하였다. 변수들 사이에 약한 선형적 연관성이 존재하는 것으로 파악되었으므로, 선형회귀모형을 이용하여 변수 간의 연관성을 검토하고 각 변수의 유의성을 평가하였다.

선형회귀모형을 통한 분석 결과, a 속도와 실압력이 탱구 두께에 미치는 영향이 통계적으로 유의하게 나타났

며, 결정계수( $R^2$ )가 0.78로 비교적 높은 설명력을 보였다. 잔차도를 분석한 결과, 오차항의 정규성 가정이 완전히 충족되지 않으며, 영향점(influential points)이 일부 존재하는 것으로 나타났다.

이는 공정 변수들이 특정 그룹을 형성하며, 극단적인 값을 가지는 경우가 존재하기 때문으로 파악된다.

분산팽창계수(VIF)를 분석한 결과, 대부분의 변수에서 4를 넘지 않는 수치로 다중공선성이 심각하지 않은 것으로 나타났다(보통 VIF가 10을 초과하면 심각한 다중공선성이 존재한다고 판단하며, 4~5 이상이면 다중공선성을 의심할 수 있다[4]). 그러나 일부 변수에서는 VIF 값이 4에 근접하여, 변수 간의 상관관계를 추가적으로 고려할 필요가 있다.

또한, 예측된 반응변수 값에서 특정 군집이 형성되는 경향이 관찰되었으며, 이는 선형회귀모형의 성능 저하를 초래할 가능성이 있다. 이러한 문제를 해결하기 위해 추가적인 변수 변환, 비선형 회귀모형 적용 또는 회귀모형을 고려하는 것이 필요할 것으로 보인다[5].

이를 기반으로 데이터 품질을 평가하는 방법론을 구축하고, 변수 간 상관관계를 정량적으로 분석하여 최적의 공정 변수 설정 기준을 도출하였다.

다변량 분석을 통해 공정 변수가 제품 품질에 미치는 정량적 영향력을 수치화하였으며, 실시간 이상 탐지를 위한 AI 기반 알고리즘을 설계하였다[6].

### III. The Proposed Scheme

#### 1. AI-driven data quality improvement

##### 데이터품질 평가 및 분석

데이터품질은 제조 공정 최적화 및 AI 기반 데이터 정제의 효과성을 판단하는 중요한 요소이다. 본 연구에서는 데이터품질의 6가지 핵심 요소를 고려하여 데이터품질 지수(DQI, Data Quality Index)를 평가하였다. 데이터품질 분석을 위해 다음 기준을 활용하였다.[7]

정확성(Accuracy)은 데이터의 이상치를 분석하여 정확성을 평가하였다. 공정 데이터에서 이상치를 감지하기 위해 Z-score를 활용하였으며, 기준값( $Z > 3$ )을 초과하는 이상 데이터의 비율을 계산하였다. 그 결과, 전체 데이터에서 이상치 비율은 3.2%로 확인되었으며, 이는 데이터 품질을 개선하기 위해 사전 정제가 필요함을 시사한다[8].

완전성(Completeness)으로 결측 데이터 비율을 분석하여 완전성을 평가하였다. 데이터셋 내에서 누락된 값의 비율을 계산한 결과, 결측 데이터 비율은 5.6%로 확인되었

다. 이 비율을 최소화하기 위해 KNN 기반 보간법과 다변량 회귀 분석을 활용하여 데이터 복원을 수행하였다.

일관성(Consistency)으로 데이터 중복성을 분석하여 일관성을 평가하였다. 중복 데이터를 제거한 후, 데이터셋 내 일관성을 측정된 결과, 중복 데이터 비율은 2.8%로 나타났다. 이는 제조 공정 데이터 수집 과정에서의 오류 가능성을 일관성 유지가 중요한 요인임을 확인할 수 있다

신뢰성(Reliability)으로 동일한 조건에서 반복 측정된 데이터의 변동성을 분석하였다. 공정 변수의 표준편차와 평균값의 비율을 계산한 결과, 신뢰성 지수는 0.87로 측정되었다.

적시성(Timeliness)으로 데이터가 최신 상태로 유지되는지 평가하였다. 본 연구에서 활용된 제조 공정 데이터는 최신 3개월 간의 생산 데이터를 포함하고 있어, 적시성 평가는 100%로 설정하였다. 실시간 제조 데이터 분석 시스템을 구축할 경우 고려한다[9].

유효성(Validity)으로 공정 변수들이 사전에 정의된 허용 범위를 충족하는지를 평가하였다. 예를 들어, 탕구 두께(c\_thickness)의 경우 사전 정의된 범위인 [21,27]을 초과하는 데이터의 비율을 분석하였다. 그 결과, 전체 데이터에서 4.3%가 허용 범위를 벗어난 것으로 나타났다. 이는 불량률과 직접적인 연관이 있다.

데이터품질 지수(DQI) 평가의 내용은 Table 1. Data Quality Assessment Summary Table로 정리했다[10].

위 6가지 요소를 종합적으로 평가하여 데이터품질 지수(DQI)를 계산한 결과, 본 연구에서 분석된 제조 공정 데이터의 평균 품질 지수는 0.91로 나타났다. 이는 AI 기반 데이터품질 개선을 적용하기 전과 후의 변화를 비교하는 기준이 될 수 있으며, 데이터품질이 제조 공정의 신뢰성 및 효율성에 미치는 영향을 분석하는 중요한 지표로 활용될 수 있다[11].

Table 2. Data Quality Assessment Summary Table

Item	Key items	Value
Accuracy	outlier value	96.8%
Completeness	missing value	94.4%
Consistency	duplicate data	97.2%
Reliability	Repeated measurement data	87.0%
Timeliness	Latest 3 month data	100%
Validity	process variable	95.7%

AI 기반 데이터 정제 및 이상 탐지 프레임워크를 적용함으로써, 데이터품질 지수를 6.5% 향상시킬 수 있음을 실험적으로 검증하였다. 이는 제조업에서 데이터 기반 품

질 관리를 적용하는 것이 실질적으로 유용한 접근 방식임을 시사한다. 향후 연구에서는 실시간 데이터품질 평가 시스템을 도입하여 보다 정교한 품질 관리 방안을 제시할 예정이다.

공정 데이터의 신뢰성을 높이기 위해 다음과 같은 AI 모델을 활용하였다.

랜덤포레스트(Random Forest) 및 오토인코더(Autoencoder) 모델을 이용하여 이상값을 감지하였다. 랜덤포레스트 모델은 500개의 트리를 사용하여 데이터 내 이상값을 감지하는 데 활용되었으며, 평균 정확도는 96.3%를 기록하였다. 또한, 오토인코더 기반 이상 탐지 기법을 적용한 결과, 민감도(Sensitivity)는 91.4%로 나타났으며, 특정 이상값을 효과적으로 제거할 수 있었다. 실험 데이터에서 이상값 제거 후 모델의 정확도는 평균적으로 4.8% 향상되었다.

그림 1의 결측 데이터 정제를 위해 KNN(K-Nearest Neighbors) 및 다변량 회귀(Multivariate Regression)를 적용하였다. KNN 기반 보간법을 활용하여 결측 데이터를 보완한 결과, 평균 정확도가 94.2%로 향상되었으며, 다변량 회귀 분석을 병행하여 오차율을 기존 7.1%에서 3.5%로 감소시켰다. 특히, KNN 기법은 이웃 개수를 5로 구현하였을 때 가장 높은 성능을 보였으며, 다변량 회귀를 적용한 경우 데이터 정합성이 기존 대비 6.8% 증가하였다. 결측값이 10% 이하로 포함된 데이터셋에서는 다변량 회귀가 더 효과적이었으나, 결측값이 15%를 초과하는 경우 KNN 보간법이 보다 높은 정확도를 제공하였다[12].

Stage	Method	Metric	Value
Outlier Detection	Random Forest (500 Trees)	Accuracy	96.3%
Outlier Detection	Autoencoder	Sensitivity	91.4%
Outlier Removal		Accuracy Improvement	4.8%
Missing Imputation	KNN (K=5)	Accuracy	94.2%
Missing Imputation	Multivariate Regression	Error Rate Reduction	7.1% to 3.5%
-	Multivariate Regression	Data Integrity	6.8% Increase
Method Selection	Missing value < 10%	Regression	-
Method Selection	Missing value > 15%	KNN	-

Fig. 3. As a result of using interpolation to compensate for missing data

#### IV. Conclusions

본 연구에서는 다양한 분류모형을 활용하여 양품과 불량인시품을 분류하는 성능을 평가하였다. 로지스틱 회귀,

의사결정나무, 랜덤포레스트, 신경망모형을 비교하였으며, 이를 시각적으로 분석하기 위해 각 모형별 예측확률의 상자그림(Boxplot)을 작성하였다.

분석 결과, 양품을 양품으로 가장 잘 예측하는 것은 의사결정나무모형이며, 반면, 불량인시품 분류 성능은 랜덤포레스트(Random Forest)가 가장 우수한 것으로 나타났다[13]. 또한, 오분류율, AUC, KS 통계량을 비교한 결과, 랜덤포레스트 모형이 불량품 분류 성능이 가장 우수한 것으로 나타났으며, 반면 로지스틱 회귀모형은 가장 낮은 성능을 보였다. 제조 공정에서 랜덤포레스트와 같은 앙상블 기법이 적용된 모형이 보다 강건한 분류 성능을 제공할 수 있음을 확인하였다[6].

제조 공정 데이터의 품질을 평가하고, AI 기반 품질 관리 기법을 활용하여 공정 최적화를 수행하였다. 실험 결과, AI 기반 데이터 정제 및 이상 탐지 기법이 공정 효율 성과 제품 품질을 대폭 증가시킬 수 있음을 확인하였다[7]. 향후 연구에서는 실시간 공정 자동화 기술과 연계한 AI 기반 품질 관리 시스템을 구축하여 더욱 효과적인 스마트팩토리 운영을 위한 방안을 제시하고자 한다.

연구에 활용된 분류모형(로지스틱 회귀, 의사결정나무, 랜덤포레스트, 신경망)은 특정 데이터셋(Oil Gasket 제조 공정)에 최적화되어 있어 타 공정 또는 산업 환경에 직접 적용 시 성능이 저하될 수 있음. 따라서 다양한 제조 공정 데이터를 활용한 추가 검증이 필요한점을 확인하였다. 특히, 최근 3개월 동안 수집된 데이터로, 장기적인 데이터 특성을 충분히 반영하지 못하여 계절성, 장기 트렌드 등의 영향이 배제되었음. 따라서 더 긴 기간의 데이터를 활용한 추가 연구가 요구되며, 오분류율, AUC, KS 통계량 등 주로 정확성 기반의 평가 지표만 사용하여 다양한 성능평가 지표(정밀도, 재현율, F1-score 등)를 고려하지 못함. 보다 포괄적인 평가를 위해 추가적인 지표를 고려할 필요가 있음을 확인하였다.

#### REFERENCES

- [1] S. Lee, H. Kim, J. Park, and Y. Choi,\*\* "Success Factors for Implementing Smart Factories in Small and Medium-Sized Enterprises," \*Journal of Industrial Integration and Management\*, vol. 5, no. 2, pp. 117-130, 2020. DOI: 10.1142/S242486222050062
- [2] C. Wang, L. Zhao, and S. Li, "Data Quality Management in Smart Manufacturing: A Review and Future Directions," IEEE Access, vol. 8, pp. 194310-194324, Oct 2020. DOI: 10.1109/ACCESS.

2020.3033445.

- [3] Y. Zhang, X. Wang, and R. Gouriveau, "A Machine Learning-Based Data Quality Improvement Framework for Manufacturing Processes," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2110-2119, March 2021. DOI: 10.1109/TII.2020.2986249.
- [4] B. Hair, J. Anderson, R. Tatham, and W. Black, "Multivariate Data Analysis," Pearson Prentice Hall, 7th edition, pp. 221-222, 2009 DOI: 10.1109/TII.2019.2957490.
- [5] G. James, D. Witten, T. Hastie, and R. Tibshirani, "An Introduction to Statistical Learning with Applications in R," Springer, 1st edition, pp. 96-100, 2013. Available at: <http://www-bcf.usc.edu/~garth/ISL/ISLR%20Seventh%20Printing.pdf>.
- [6] W. Li, K. Zhang, and J. Lv, "Data-Driven Quality Prediction and Anomaly Detection for Manufacturing Processes Based on Deep Learning," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3700-3709, June 2020. DOI: 10.1109/TII.2019.2945430.
- [7] S. Bardi, M. Sorli, and G. Toschi, "A Methodology for Data Quality Assessment in Manufacturing Systems," *International Journal of Advanced Manufacturing Technology*, vol. 92, no. 1-4, pp. 117-130, Sept 2017. DOI: 10.1007/s00170-017-0354-7.
- [8] Hadi, and M. Nyitrai, "Outlier Detection and Data Cleaning in Manufacturing," *Procedia CIRP*, vol. 55, pp. 287-292, 2016. DOI: 10.1016/j.procir.2016.07.043
- [9] L. Breivold, and I. Sandkuhl, "Data Quality in Real-Time Manufacturing Control," *Computers in Industry*, vol. 61, no. 7, pp. 678-687, Sept 2010. DOI: 10.1016/j.compind.2010.02.013
- [10] R. Rahim, M. Husin, and N. Haron, "Data Quality Assessment for Manufacturing Processes," *Journal of Physics: Conference Series*, vol. 1019, no. 1, pp. 012014, May 2018. DOI: 10.1088/1742-6596/1019/1/012014
- [11] M. Sorli, S. Bardi, and G. Toschi, "A Comprehensive Data Quality Assessment for Manufacturing Process Improvement," *International Journal of Production Research*, vol. 56, no. 1-2, pp. 679-694, Jan 2018. DOI: 10.1080/00207543.2017.1369502
- [12] M. Garcia-Laencina, J. Sancho-Gomez, J. A. Lozano, and P. Larranaga, "A Review of Imputation Methods on Missing Values," *Applied Intelligence*, vol. 33, no. 1, pp. 22-41, Aug 2010. DOI: 10.1007/s10489-009-0228-z
- [13] S. Wuest, D. Weimer, C. Irgens, and K.-D. Thoben, "Machine Learning in Manufacturing: Advantages, Challenges, and Applications," *Production Engineering*, vol. 9, no. 6, pp. 731-748, Dec 2015. DOI: 10.1007/s11740-015-0636-9

## Authors



Heung-Sup Sim received the B.S. Information and Communication Engineering, received the M.S. degrees in advertising science from Chung-Ang University, Korea, in 2010, M.S. and Ph.D. completion Dongyang University, Korea in 2013.

He is currently a Professor in Dept. of Digital Twin Software Dongyang University. He is currently a professor of computer and military science at Dongyang University. He is interested in ESG data analysis, smart farm advancement, and IOT-cloud computing.