

반도체 제조공정 변수 중요도 도출에 대한 연구

최은선*, 전유정*, 김진실*, 김재성**, 조완섭**
충북대학교 빅데이터융합과정*, 충북대학교 경영정보학과**

{tmxk147, ddbb0147, kim94203069}@gmail.com*, comkjsb@cbnu.ac.kr**

A Study on Deriving the Importance of Variables Affecting the Semiconductor Manufacturing Process

Eun-Seon Choi*, Yu-Jung Janu*, Jin-Sil Kim*, Jae-Sung Kim**, Wan-Sup Cho**
Chung-Buk National Univ.

요약

스마트팩토리는 ICT 기술을 전체 생산 공정에 적용해 실시간으로 수집된 대량의 데이터를 분석·관리할 수 있는 지능형 생산공장으로 생산성과 효율성을 높일 수 있다는 점에서 제조업계의 큰 이슈로 떠오르고 있다. 이에 따라 반도체 시장에서도 스마트 공장에 대한 연구가 진행 중이다. 이때, 반도체 제조에 중요한 것은 향후 제품 수량 계획과 같은 사업 의사결정에 큰 영향을 미치는 반도체 수율과 불량률을 예측하는 것이다. 따라서 제조 공정 과정에서 수집된 다양한 데이터 중에 해당 반도체 수율 및 불량률에 영향을 크게 미치는 변수가 무엇인지 파악되어야 한다. 본 연구에서는 반도체 생산의 기초 소재인 웨이퍼의 생산공정에서 수집된 데이터를 통해 불량에 영향을 미치는 핵심적인 변수를 SHAP 로 도출하여 후행 연구로 진행될 예측의 기반을 다지고자 한다.

I. 서론

스마트팩토리에서 기존 기계와 ICT 융합은 기계의 고부가가치화를 이루고, 공정 과정의 다양한 제조 데이터 수집 및 빅데이터 분석을 통해 다양한 공정 및 수준별 공장에 적용 가능한 생산 설비의 자율 교정 및 자가 조직화를 가능하게 한다(서창성, 정진진, 김석찬, 2018). 반도체 제조공정 또한 시스템의 자동화로 하루에도 수백, 수천만 개의 데이터가 생성되고 데이터베이스화되어 실시간 공정 상태를 파악할 수 있다. 반도체 제조업체들이 글로벌 경쟁력 향상 및 수익성을 유지하기 위해서는 공정 과정에서 생성된 데이터를 바탕으로 높은 수율 유지, 불량품 처리 및 역동적인 제조 환경을 고려한 프로세스 개선 등 다양한 문제를 해결해 나가야 하며 이를 위해 노력하고 있다. 효과적이고 효율적으로 데이터를 분석하기 위해서는 필요한 변수만 선택하여 사용해야 한다. 모든 변수를 고려하여 분석하면 분석 시간이 길어져 빠른 의사결정이 불가능하다. 따라서, 분석 과정에서 중요 요인을 미리 파악해야 하고 제조 및 공정 과정에서 발생하는 수많은 변수 중에 유의미한 변수를 도출할 수 있어야 한다.

II. 본론

본 연구에서는 R 을 사용하여 의사결정나무와 XGBoost, 그리고 SHAP 기법을 통해 불량에 영향을 많이 미치는 변수의 중요도를 도출하는 것을 목표로 한다. 의사결정나무는 매우 간단한 논리로 표현되므로 실무에 적용하기 편리하고 분석 방법과 결과를 직관적으로 이해하기 쉽다는 장점이 있기에 (이민수, 최영찬, 유병준, 2007), 모형 훈련과 조절 과정을 간소화하는 함수를 포함하는 caret 패키지를 사용하였다. 또한, XGBoost 와 SHAP 를 추가로 사용하여 변수

중요도를 도출하고 이를 비교하였다. XGBoost 는 머신러닝 모델 사이에서 예측 성능이 뛰어나다고 알려진 알고리즘으로, 트리 모델을 병렬로 운용하기 때문에 빠르게 학습하고 분류할 수 있다는 장점이 있으며, SHAP 는 머신러닝 모델에서 각 기능의 중요성과 설명을 계산하기 위해 응용하는 인기 있는 기술 중 하나이다 (최은선, Som Akhamixay Oui, 전유정, 김진실, 2020).

사용 데이터는 반도체 부품을 생산하는 A 공장에서 생성된 데이터이다. 데이터에는 반도체 생산과 관련된 다양한 28 개의 변수가 존재한다.

Table 1. 사용변수 설명

그룹	변수
A	X1~ X5
B	X6~X10
C	X11~X15
D	X16~X20
E	X21~X24
F	X25~X28

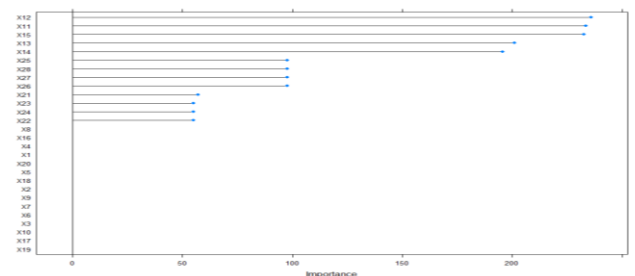


그림. 1. Feature Importance of Test (caret)

<그림 1>은 caret 패키지를 사용하여 도출한 변수 중요도를 시각화한 것이며, 아래 표는 caret 패키지를 사용하여 도출한 변수 중요도 중 상위 5 개를 나타낸 표이다. 변수 중요도 도출 결과 X12, X11, X15, X13, X14 순으로 중요도가 높은 것으로 확인되었다.

Table 2. Importance of Each Variable (caret)

Feature	Gain
X12	236.0235
X11	233.5766
X15	232.5004
X13	201.0407
X14	195.7607
X12	236.0235

<그림 2>는 XGBoost 를 사용하여 도출한 변수 중요도를 시각화한 것이다.

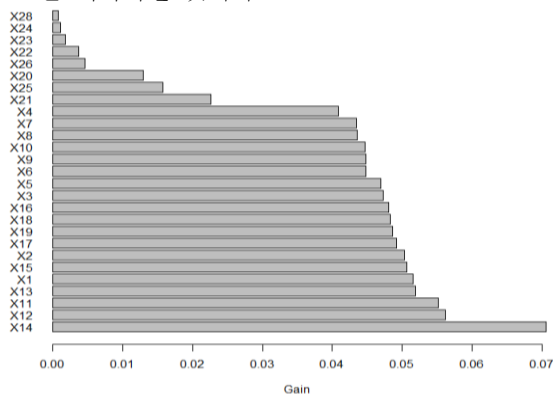


그림. 2. Feature Importance of Test (XGBoost)

아래 표는 XGBoost 를 사용하여 도출한 변수 중요도 중 상위 5 개를 나타낸 표이다. 변수 중요도 도출 결과 X14, X12, X11, X13, X1, X15 순으로 중요도가 높은 것으로 확인되었다.

Table 3. Importance of Each Variable (XGBoost)

Feature	Gain
X14	0.070667
X12	0.05617
X11	0.055167
X13	0.05187
X1	0.051525
X15	0.05064

<그림 3>은 SAHP 를 사용하여 도출한 변수 중요도를 시각화 한 것이다.

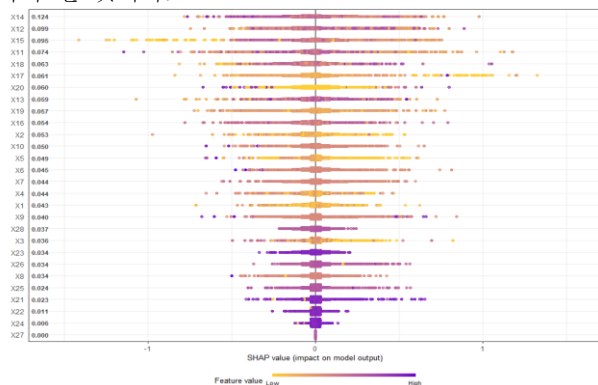


그림. 3. Feature Importance of Test (SHAP)

아래 표는 SHAP 를 사용하여 도출한 변수 중요도 중 상위 5 개를 나타낸 표이다. 변수 중요도 도출 결과 X12, X11, X15, X13, X14 순으로 중요도가 높은 것으로 확인되었다.

Table 4. Importance of Each Variable (SHAP)

Feature	Gain
X14	0.124081
X12	0.098702
X15	0.095189
X11	0.073679
X18	0.063408
X14	0.124081

III. 결론

본 논문에서는 측정된 데이터를 바탕으로 불량률에 영향을 미치는 중요 변수를 알아보기 위해 의사결정나무와 XGBoost, 그리고 SHAP 를 사용하여 변수 중요도를 도출하였다. 그 결과 각 모델에서 도출한 변수 중요도 대부분이 <표 1>에서 정의한 그룹 C 에 속하는 변수임을 확인할 수 있었으며, 그룹 C 환경의 각별한 관리의 필요성이 대두된다. 또한 공정 과정에서 발생하는 불량 유형이 다양하고 이에 영향을 미치는 미지의 변수도 존재하기 때문에 생산에 영향을 미치는 미지의 중요 변수를 밝혀낼 필요가 있다. 변수의 증가에 따른 효율적인 변수 선택 방법 및 새로운 학습 방법이 필요하기에 추후 연구에서는 본 연구 결과를 바탕으로 더 발전된 알고리즘을 사용하여 모델의 분류 및 예측 정확도를 높이는 연구를 진행할 것이다.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (Project ID:P0011880).

참 고 문 헌

서창성, 정신진, 김석찬 (2018). 기업의 생산성 향상을 위한 스마트 팩토리 구축. 한국통신학회지(정보와통신), 35(6), 43-49

이민수, 최영찬, 유병준 (2007). 의사결정나무법을 이용한 귀납적 학습방법에 의한 정보시스템 수용자 세분화. Information Systems Review, 9(1), 67-84.

최은선, Som Akhamixay Oui, 전유정, 김진실 (2020). A Study on Deriving the Importance of Variables Using XGboost and SHAP.