딥러닝을 활용한 반도체 제조 공정의 주요 인자 분석
A Study on the Factor Analysis in Semiconductor Manufacturing Process

신윤경, 길준혁, 이나영, 배장원
한국기술교육대학교 산업경영학부
{rt0326, gjhkms00929, 4yeong0318, jangwon_bae}@koreatech.ac.kr

Abstract 4차 산업혁명 이후 빅데이터는 제조 현장 대부분의 과정에서 활용 및 처리에 대한 중요성이 부각되고 있다. 본 논문에서는 제조 수율 개선을 위해 반도체 제작 공정 데이터를 분석한다. 공정 데이터는 반도체 공정에서 실시간으로 확보되는 분석에 부적합한 형태의 데이터에서 주요 인자 분석에 적합한 형태로 전처리가 완료된 데이터를 이용한다. 전처리된 데이터를 회귀 분석하고, 딥러닝 기반의 예측 모델링을 구축하여 반도체 제조 공정에서 불량률을 낮추는데 중요한 인자를 분석하였다. 예측 모델의 성능 판단을 위해 각각 MAPE와 F1 score를 측정하였다.

1. 서론

빅데이터는 기존 데이터보다 너무 방대하여 기존의 방법이나 도구로는 수집/저장/분석 등이 어려운 정형 및 비정형 데이터들을 의미한다. 4차 산업혁명 이후 빅데이터의 가치 및 중요성은 더욱 부각되었으며, 세계적인 컨설팅 기관인 매켄지(Mckinsey)는 빅데이터를 기존 데이터베이스 관리도구의 데이터 수집, 저장, 관리, 분석하는 역량을 넘어서는 규모로서 그 정의는 주관적이며 앞으로도 계속 변화할 것이라 언급하였다[1].

제조 산업에서 역시 빅데이터는 큰 주목을 받고 있는데, 국내외 제조 현장에서 이를 활용해 스마트팩토리를 구축하는 사례들이 증가하고 있다. 제조 및 품질 데이터를 토대로 불량품을 실시간으로 감지하고 불량품의 추가 생산을 방지하여 제품의 완성도는 높이고 불량률은 낮추는 것이다. 빅데이터 분석을 통한 공정 효율화는 생산성을 향상시킴과 동시에 이를 토대로 제품의 수요 및 시장의 변화에 빠르게 대응할 수 있도록 한다.

빅데이터의 중요성이 대두되며 함께 주목받기 시작한 기술이 바로 딥러닝이다. 기존에 데이터간의 관계를 분석하고 모델식을 만드는데 사용된 회귀분석(Regression Analysis)은 간단하고 쉽게 결과를 도출할 수 있다는 장점이 있지만, 정확성이 상대적으로 떨어지며 모델의 정교화를 위해서는 분석가의 노력을 요구한다는 단점을 지니고 있었다. 그래서 등장한 것이 인공신경망(Artificial Neural Network, ANN)이며, 여기에 빅데이터를 결합한 것을 딥러닝(Deep Learning)이라 한다[2].

그림 1 딥러닝, 빅데이터 관계도

딥러닝은 컴퓨터가 스스로 외부 데이터를 조합, 분석하여 학습하는 기술을 뜻한다. 축적된 데이터를 단순히 분석만 하지 않고 이 데이터를 통해 학습까지 하는 기계 학습 능력을 활용해 최적의 결론을 내린다[3]. 최근 많은 제조 기업이 기존 규칙 기반 검사 시스템을 보완하고 생산라인의 효율을 높이기 위하여 딥러닝 기반 검사 시스템 도입을 고려하고 있

다. 딥러닝 기반 검사 시스템은 기존 규칙 기반 검사로 확인이 쉽지 않아 수작업으로 해야만 했던 업무를 대체하여 인건비를 줄이고, 생산의 효율을 증대시킨다. 또한, 신속한 구현이 가능하며, 검사 결과를 문서화하는 딥러닝 솔루션을 이용한다면, 새로운 오류가 발생할 경우 이전의 학습 데이터를 통해 보완이 필요한 점을 쉽게 파악하여 결함을 신속하게 확인 및 제거할 수 있도록 한다[4].

세계적으로 제조업 분야에서 딥러닝을 기반으로 한 스마트팩토리는 각광을 받고 있으나, 국내 중소기업 환경에서의 구체적 활동 전략이나 성공사례는 많지 않다[5].

특히 정밀 규격공차가 요구되는 반도체 제작 공정에서는 공정 과정에서 확보한 빅데이터를 통해 주요 인자를 분석하고 품질을 향상시키고자 하는 필요성이 부각되고 있다. 기존에 사용되던 회귀 분석 및 요인 분석은 기본 가정이 어긋날 경우 사용이 불가하며, 분석 결과의 신뢰성 문제가 발생하는 등의 단점이 존재했다. 따라서 이러한 문제들을 개선하기 위해 빅데이터 기반의 공정 요인 분석이 필요하다.

본 논문에서는 제조 수율 개선을 위해 반도체 제작 공정 데이터를 분석한다. 공정 데이터는 반도체 공정에서 실시간으로 확보되는 분석에 부적합한 형태의 데이터에서 주요 인자 분석에 적합한 형태로 전처리가 완료된 데이터를 이용한다. 전처리된 데이터를 회귀 분석하고, 딥러닝 기반의 예측 모델링을 구축하여 반도체 제조 공정에서 불량률을 낮춰주는데 중요한 인자를 분석하였다. 이후 예측 모델의 성능 판단을 위해 각각 MAPE와 F1 score를 측정하였다. 본 논문에서 분석을 통해 얻은 결과는 현장 작업자들에게 제작 공정의 주요 인자 정보를 제공해 줄 수 있으며 향후 해당 정보를 기반으로 품질 향상을 위한 제안을 도출할 수 있을 것으로 전망한다.

본 논문은 1장 서론에 이어, 2장에서 회귀분석 기법 및 딥러닝 모델에 대해 서술한다. 3장에서는 회귀분석과 딥러닝 기반 공정 데이터 분석을 한다. 4장에서는 MAPE와 F1 score를 통한 모델 평가를 하며 5장은 결론으로 구성한다.

2. 관련 연구

2.1 로지스틱회귀분석(logistic regression analysis)

회귀분석(regression analysis)은 매개변수 모델(parametric model)을 이용하여 통계적으로 변수들 사이의 관계를 추정하는 분석방법이다. 주로 독립변수(independent variable)가 종속변수(dependent variable)에 미치는 영향을 확인하고자 사용한다. 회귀분석은 다른 독립변수들을 고정시키고 한 가지 독립변수만을 변화시킬 때 종속변수가 어떻게 변화하는지를 확인한다. 종속변수와 관련이 있는 독립변수를 찾을 때나 독립변수들 간의 관계를 이해하고자 할 때 사용한다[6].

로지스틱회귀는 종속 변수가 이분형일 때 수행할 수 있는 회귀 분석 기법의 한 종류이다. 로지스틱 회귀는 하나의 종속 이진 변수와 하나 이상의 숫자형, 명목형, 순서형의 독립변수 간의 관계를 로지스틱 회귀함수를 이용하여 정량적으로 설명한다[7].

본 논문에서는 딥러닝 기반 분석 결과와의 비교를 위해 분량 판정을 True, False로 구분하여 분석한 로지스틱 모델을 활용하였다.

2.2 딥러닝(deep learning) 모델

딥러닝은 머신러닝의 여러 방법론 중에서 인공신경망을 여러 층 쌓아 올린 기법을 의미하는데, 인공신경망을 깊게 쌓아 올려 학습하기 때문에 딥러닝이라 불린다. 딥러닝 알고리즘 근간인 인공신경망 기법은 1980년대부터 활발히 연구되어 다양한 산업군에 사용되어왔으며, 최근에는 딥러닝 알고리즘이 주목을 받고 있다. 본 논문에서 사용된 공정 데이터와 같은 대량의 데이터를 다루는 데에 딥러닝 알고리즘이 탁월한 역할을 하며, 이는 기존의 머신러닝 기법 등 인공지능 기법들을 월등히 뛰어넘는 성능을 보여준다. 딥러닝 알고리즘은 인공신경망을 깊게 쌓아 올리는데, 이로써 데이터의 특징을 단계별로 학습할 수 있다는 효과를 얻는다. 이처럼 데이터의 특징을 단계별로 학습하기 때문에 딥러닝을 표현 학습이라고도 부른다. 데이터의 특징을 잘 학습하면 학습한 특징을 이용하여 알고리즘이 보다 좋은 성능을 도출할 수 있을 것으로 보인다.

본 연구에 데이터를 제공한 'H'사에서는 도넛 모양의 웨이퍼를 제작하는데, 45°간격으로 점이 찍히며, 이는 각각 기울기를 의미한다. 그리고 이때 각 지점 및 웨이퍼의 두께를 PCD라고 칭한다. PCD는 1차 가공 후 한 번, 2차 가공 후 한 번 총 두 번 측정하는데, 1차와 2차 PCD 간 차이를 계산한 후 도출된 값의 최솟값과 최댓값을 각각 구해 PCD_difference_Min_1, PCD_difference_Min_2, PCD_difference_Max_1, PCD_difference_Max_2로 저장

해둔다. 이후 1차와 2차 PCD difference_Min(_1 혹은 _2) 칼럼의 'total' 인덱스 값 + PCD difference_Max(1 혹은 _2) 칼럼의 'total' 인덱스 값을 각각 slope_1과 slope_2라 한다. slope_1과 slope_2의 차이는 slope_change라 한다. 해당 기업에서는 1차 가공이 끝난 후 slope값의 범위에 따라 'tilting'이라는 값을 설정한다. 해당 값들을 계산하여 Tilt_table을 작성하였는데, 해당 Tilt_table과 비교하여 어떤 Tilting 값을 설정하였는지 나타내는 칼럼으로 Tilt_judgement를 도출하였다. 그리고 제품을 측정했을 때 pass인지 unpass인지 확인하여 unpass 값이 하나라도 존재할 경우 False, 모두 pass일 경우True로 판단한 값들을 Tilt_spec으로 두었다.

본 논문에서는 딥러닝 모델을 활용하여 (1) 'slope_1', 'slope_2', 'slope_change'와 'Tilt_judgement' 간, (2) 'slope_1', 'slope_2', 'slope_change', 'Tilt_judgement'와 'Tilt_spec' 간 연관 관계를 파악하는 것을 목적으로 한다. (1)번의 관계를 분석한 것이 'DNN_1', (2)번의 관계를 분석한 것이 'DNN_2'이다. 이를 통해 회귀 모델에서는 나타나지 않았던 심층 관계를 파악하여 반도체 제조 공정의 주요 인자를 확인하고자 한다.

3. 반도체 공정의 주요 인자 분석

3.1 반도체 공정 측정 데이터

본 논문의 연구를 위해 데이터를 제공해 준 국내 제조업체 H사의 웨이퍼 두께 가공 공정은 그림 2와 같이 나타내었다. 웨이퍼 가공을 위해 크게 'Cropping', 'Slicing', 'Surface Grinding1', 'Surface Grinding2'의 과정을 거치는데, 본 논문에서는 마지막 단계의 표면 기울기를 관리하는 것을 목표로 한다. 회귀모델과 딥러닝 모델에 사용된 Product code는 표 1과 같다. 선형회귀와 로지스틱회귀, 'DNN_2'에는 A3R001-SM(데이터 3,099개), A3R002-SM(데이터 2,578개), A3R003-SM(데이터 510개), A3R004-SM(데이터 2,756개)의 4개 제품이 모두 사용되었으며, 'DNN_1'에는 A3R001-SM(데이터 3,099개), A3R003-SM(데이터 510개)의 2개 제품이 사용되었다. Train data와 Test data의 비율은 모두 8:2로 동일하게 설정되었다.

Product_code	Train : Test = 8 : 2	
A3R001-SM (3,099)	Train	2,479
	Test	620
A3R002-SM (2,578)	Train	2,062
	Test	516
A3R003-SM (510)	Train	408
	Test	102
A3R004-SM (2,756)	Train	2,205
	Test	551

표 1 Train, Test 데이터의 비율 및 수

3.2 로지스틱회귀 모델 분석

불량 판정을 '불량이다, 아니다'로 나타내기 위해 로지스틱 회귀 모델을 이용했다. 로지스틱회귀 모델 구축은 Python을 사용하였으며, x, y 변수를 설정하여 학습을 진행한 후 로지스틱회귀 분석을 시행하였다. Y값은 True, False 또는 0, 1로 도출시키기 위해 모두 'Tilt_spec'으로 설정하였으며, X값은 다음과 같이 설정하였다(표 2 참고).

X값
slope_change, tilt_judgement
slope_1, tilt_judgement
slope_2, tilt_judgement
slope_1, slope_change, tilt_judgement
slope_1, slope_2, slope_change, tilt_judgement

표 2 로지스틱회귀 분석에 사용된 X값

이후 최적의 변수를 찾기 위해 전진 선택법, 후진 소거법, 전진 단계적 선택법의 변수 선택법을 활용했다.

3.3 딥러닝 기반 공정 데이터 분석

3.3.1 DNN 1

[DNN_1]에서 역시 모델 구축을 위해 Python을 사용하였다. 먼저 난수의 시드값을 지정한 후, 기본 하이퍼파라미터를 지정한다. 이때, optimizer에 지정할 학습률을 의미하는 learning_rate는 0.05로 지정하고, 전체

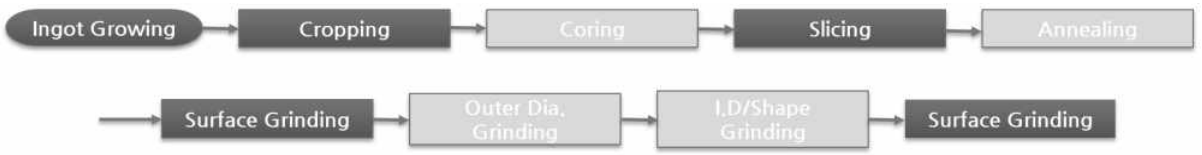

그림 2 두께 가공 공정

데이터셋을 몇 번 반복해서 학습할 지를 정하는 epoch 수는 100으로 지정하였으며, 파라미터를 데이터 몇 개 당으로 나눌 것인지를 정하는 batch_size는 30으로 지정하였다. 데이터는 역시 각 csv 파일에서 로드하였으며, x와 y로 변수를 지정하여 x_train, x_test, y_train, y_test로 나누었는데, y값은 모두 tilt_judgement로 설정하였다. 각각 설정된 x값은 표 3과 같다.

x값
slope_change
slope_1
slope_2
slope_1, slope_2
slope_1, slope_change
slope_2, slope_change
slope_1, slope_2, slope_change

표 3 DNN_1 분석에 사용된 x값

이후, 데이터셋으로 변환하여, 모델을 생성하고 학습 후 최종 검증을 통해 예측값을 도출하고 이에 대한 loss[1] 체크를 진행하였다. 해당 모델에서 각 'slope 1, 'slope 2', 'slope_change' 데이터는 모델의 입력층에 활용되며, 'Tilt_judgement' 데이터는 입력된 데이터에 따른 값을 도출하기 위해 학습 시 활용된다. 따라서 데이터의 특징상 입력층에 활용된 독립변수 데이터들을 통해 'Tilt_judgement' 변수와의 관계를 확인할 수 있도록 학습을 진행한다.

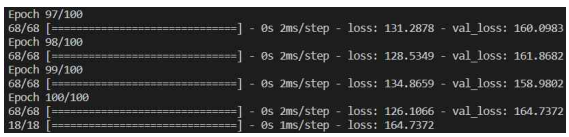

그림 3 DNN_1 학습 결과 예시

3.3.2 DNN_2

[DNN_2] 역시 모델 구축을 위해 Python을 사용하였으며, train과 test 데이터의 비율을 8:2로 준비하였다. 이때 random_state는 7로 지정하였다. 해당 모델은 'Tilt_spec'의 이진 분류를 위해 sigmoid를 사용하여 모델을 설계하였으며, 훈련 과정에서 모델 성능 향상이 없을 경우 조기 종료를 하기 위해 EarlyStopping을 활용하였다. 이후 평가 후 예측값을 도출할 수 있도록 하였으며, 이 역시 loss 체크를 진행하였다. 이때, y값은 모두 Tilt_spec으로 설정하였으며, 각각 설정된 x값은 표 4와 같다.

x값
Tilt_judgement
slope_1
slope_2
slope_change
slope_1, slope_2
slope_1, slope_change
slope_1, Tilt_judgement
slope_2, slope_change
slope_2, Tilt_judgement
slope_change, Tilt_judgement
slope_1, slope_2, slope_change
slope_1, slope_2, Tilt_judgement
slope_1, slope_change, Tilt_judgement
slope_2, slope_change, Tilt_judgement
slope_1, slope_2, slope_change, Tilt_judgement

표 4 DNN_2 분석에 사용된 x값

[DNN_2]에서 각 'slope_1', 'slope_2', 'slope_change', 'Tilt_judgement' 데이터가 모델의 입력층에 활용되며, 'Tilt_spec'의 데이터가 입력된 데이터에 따른 값을 도출하기 위해 학습 시 활용된다. 따라서 데이터의 특징상 입력층에 활용된 독립변수 데이터들을 통해 'Tilt_spec'변수와의 관계를 확인할 수 있도록 학습을 진행한다.

4. 모델 평가

4.1 MAPE - DNN_1

DNN_1 모델은 성능 판단을 위해 MAPE(Mean Absolute Percentage Error)를 사용하였다. MAPE의 계산 식은 아래와 같다.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{f}(x_i)}{y_i}\right| \quad (y_i \neq 0)$$

평가 결과 A3R001-SM 제품은 0.007로 MAPE 값이 가장 작게 나온 Slope_2를 대입했을 때가 최적인 것으로 나타났으며, A3R003-SM 제품은 0.036으로 MAPE 값이 가장 나온 Slope_change를 대입했을 때가 최적인 것으로 나타났다. 해당 결과는 표 5와 같다.

1) loss : 훈련 손실값. val_loss : 검증 손실값. accuracy : 훈련 정확도, val_accuracy : 검증 정확도

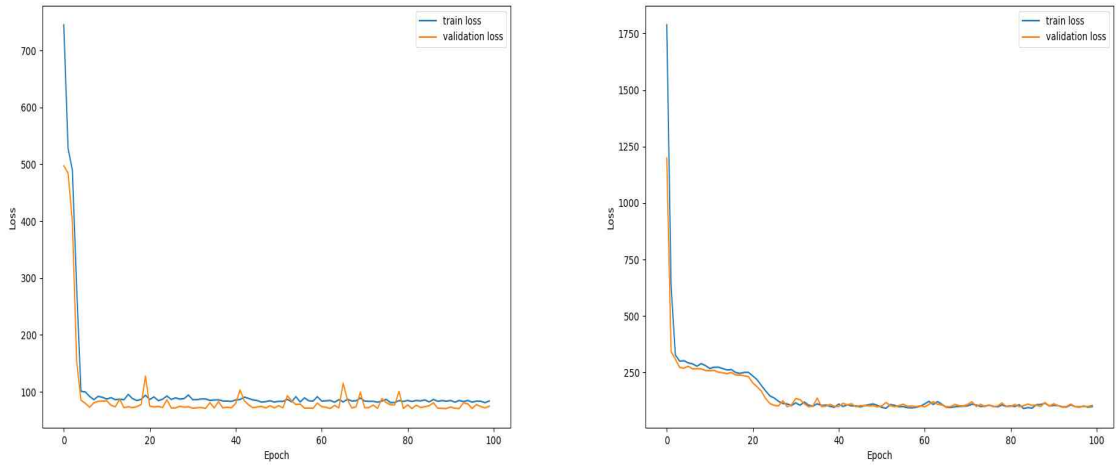

그림 4 DNN 1 모델 성능 평가 예시

제품코드	Slope_change	Slope_1	Slope_2	Slope_1, Slope_2	Slope_1, Slope_change	Slope_2, Slope_change	Slope_1, Slope_2, Slope_change
A3R001-SM	0.013	0.056	0.007	0.053	0.027	0.054	0.58
A3R003-SM	0.036	0.155	0.229	0.157	0.183.	0.142	0.191

표 5 DNN_1 모델 평가 결과

4.2 F1 score

4.2.1 로지스틱회귀 모델

로지스틱 모델의 성능을 평가하기 위해서는 Confusion Matrix를 이용했다. 예측값과 실측값은 TP(True Positive), FN(False Negative), FP(False Positive), TN(True Negative)의 행렬을 갖는다. 이때, TP는 실제 True인 정답을 True라고 예측한 경우, FP는 실제 False인 정답을 True라고 예측한 경우, FN은 실제 True인 정답을 False라고 예측한 경우, TN은 실제 False인 정답을 False라고 예측한 경우를 의미한다. 이후, 재현율, 정밀도, F1 score를 이용하여 추가 성능 평가를 진행하였다. 재현율은 실제값이 Positive인 대상 중에 예측값과 실제값이 Psitive로 일치한 데이터의 비율을 의미하며, 정밀도는 예측을 Positive로 한 대상 중에 예측값과 실제값이 Positive로 일치한 데이터의 비율을 의미한다. F1 score는 정밀도와 재현율의 조화 평균으로 해석할 수 있다.

평가 결과 A3R001-SM 제품은 0.945로 F1 score 값이 가장 크게 나온 slope_1, slope_2, slope_chage, thickness_min_1, thickness_avg_1, undulation_1를 대입했을 때 최적인 것으로 나타났다. A3R002-SM 제품은 0.928로 F1 score 값이 가장 크게 나온 thickness_min_2을 대입했을 때 최적인 것으로 나타났고, A3R003-SM 제품은 0.985로 F1 score 값이 가장 크게 나온 thickness_min_2를 대입했을 때 최적인 것으로 나타났으며, A3R004-SM 제품은 0.977로 F1 score 값이 가장 크게 나온 thickness min 2을 대입했을 때 최적인 것으로 나타났다. 해당 결과는 표 6과 같다.

	precision	recall	f1-score	support
False	0.82	0.49	0.62	93
True	0.92	0.98	0.95	527
accuracy			0.91	620
macro avg	0.87	0.74	0.78	620
weighted avg	0.90	0.91	0.90	620

그림 5 로지스틱회귀 모델 성능 평가 예시

제품코드	F1_score	변수 선택법;x값
A3R001-SM	0.945	후진 소거법;slope_1, slope_2, slope_change, thickness_min_1, thickness_avg_1, undulation_1
A3R002-SM	0.928	전진 단계별 선택법;thickness_min_2
A3R003-SM	0.985	전진 단계별 선택법;thickness_min_2
A3R004-SM	0.977	전진 선택법;thickness_min_2

표 6 로지스틱회귀 모델 평가 결과

제품코드	Tilt_judgement	Slope_1	Slope_2	Slope_change	Slope_1, Slope_2	Slope_1, Slope_change	Slope_1, Titl_judgement
A3R001-SM	0.839	0.838	0.898	0.839	0.898	0.892	0.837
A3R002-SM	0.837	0.837	0.837	0.837	0.837	0.837	0.837
A3R003-SM	0.941	0.941	0.941	0.941	0.941	0.941	0.941
A3R004-SM	0.947	0.947	0.947	0.947	0.947	0.947	0.947

제품코드	Slope_2, Slope_change	Slope_2, Tilt_judgement	Slope_change, Tilt_judgement	Slope_1, Slope_2, Slope__change	Slope_1, Slope_2, Tilt_judgement	Slope_1, Slope_change Tilt_Judgement	Slope_2, Slope_change, Tilt_judgement	Slope_1 Slope_2, Slope_change, Tilt_judgement
A3R001-SM	0.899	0.837	0.839	0.912	0.839	0.839	0.839	0.839
A3R002-SM	0.837	0.837	0.837	0.837	0.837	0.837	0.837	0.835
A3R003-SM	0.941	0.941	0.941	0.941	0.941	0.941	0.941	0.941
A3R004-SM	0.947	0.947	0.947	0.947	0.952	0.947	0.947	0.947

표 7 DNN_2 모델 평가 결과(precision_score)

제품코드	Tilt_judgement	Slope_1	Slope_2	Slope_change	Slope_1, Slope_2	Slope_1, Slope_change	Slope_1, Titl_judgement
A3R001-SM	1.0	1.0	0.994	1.0	0.994	0.998	0.990
A3R002-SM	1.0	1.0	1.0	1.0	1.0	1.0	1.0
A3R003-SM	1.0	1.0	1.0	1.0	1.0	1.0	1.0
A3R004-SM	1.0	1.0	1.0	1.0	1.0	1.0	1.0

제품코드	Slope_2, Slope_change	Slope_2, Tilt_judgement	Slope_change, Tilt_judgement	Slope_1, Slope_2, Slope__change	Slope_1, Slope_2, Tilt_judgement	Slope_1, Slope_change Tilt_Judgement	Slope_2, Slope_change, Tilt_judgement	Slope_1 Slope_2, Slope_change, Tilt_judgement
A3R001-SM	0.994	0.990	1.0	0.979	1.0	1.0	1.0	1.0
A3R002-SM	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.984
A3R003-SM	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
A3R004-SM	1.0	1.0	1.0	1.0	0.994	1.0	1.0	1.0

표 8 DNN_2 모델 평가 결과(recall_score)

제품코드	Tilt_judgement	Slope_1	Slope_2	Slope_change	Slope_1, Slope_2	Slope_1, Slope_change	Slope_1, Titl_judgement
A3R001-SM	0.912	0.912	0.943	0.912	0.943	0.942	0.907
A3R002-SM	0.991	0.911	0.911	0.911	0.911	0.911	0.911
A3R003-SM	0.941	0.967	0.970	0.970	0.970	0.967	0.970
A3R004-SM	0.973	0.973	0.973	0.973	0.973	0.973	0.973

제품코드	Slope_2, Slope_change	Slope_2, Tilt_judgement	Slope_change, Tilt_judgement	Slope_1, Slope_2, Slope__change	Slope_1, Slope_2, Tilt_judgement	Slope_1, Slope_change Tilt_Judgement	Slope_2, Slope_change, Tilt_judgement	Slope_1 Slope_2, Slope_change, Tilt_judgement
A3R001-SM	0.944	0.907	0.912	0.944	0.912	0.912	0.912	0.912
A3R002-SM	0.911	0.911	0.911	0.911	0.911	0.911	0.911	0.903
A3R003-SM	0.970	0.970	0.970	0.970	0.970	0.970	0.970	0.970
A3R004-SM	0.973	0.973	0.973	0.973	0.973	0.973	0.973	0.973

표 9 DNN_2 모델 평가 결과(F1 score)

4.2.2 DNN_2

DNN_2 모델은 성능 판단을 위해 정밀도와 재현율 및 이들의 조화 평균 값이 F1 score를 사용하였다. 평가 결과 A3R001-SM 제품은 0.944로 F1 score 값이 가장 크게 나온 Slope_2, Slope_change와 Slope_1, Slope_2, Slope_change를 대입했을 때가 최적인 것으로

나타났다. A3R002-SM 제품은 0.991로 F1 score 값이 가장 크게 나온 Tilt judgement를 대입했을 때가 최적인 것으로 나타났고, A3R003-SM 제품은 Tilt judgement와 Slope_1와 Slope_1, Slope_change를 대입했을 경우를 제외한 모든 결과가 0.970으로 동일하게 높은 수치를 기록해 유의미한 결과를 보이지 않았다. 마지막으로 A3R004-SM 제품 역시 모든 경우의 결과 값이 동일해 유의미한 결과를 보이지 않았다. 해당 결과는 표 8과 같다.

5. 결론

본 논문에서는 정밀 규격공차가 요구되는 제작 공정의 주요 인자 분석을 진행하였다. 공정 데이터는 반도체 공정에서 실시간으로 확보되는 분석에 부적합한 형태의 데이터에서 주요 인자 분석에 적합한 형태로 전처리가 완료된 데이터를 이용하였다. 전처리 된 데이터를 이용해 공정 변동 요인 분석을 위하여 선형 회귀 분석과 로지스틱 분석 및 2가지의 딥러닝 기반 분석을 진행하였다. 각 분석 모델의 성능을 확인하기 위해 제품 및 변수 별로 MAPE와 F1 score를 측정한 결과 제품과 모델 별로 최적값이 상이하게 나타났다. 하지만 기본적으로 딥러닝 기반 분석의 경우 MAPE 값은 낮은 수치를 기록하였으며, F1 score 값은 높은 수치를 달성하였으므로 더 많은 제품들의 데이터가 확보되면 충분히 유의미한 결과를 보일 것으로 예상된다. 이러한 회귀 및 딥러닝 기반 분석들은 효과적인 공정 변동 감소를 통한 품질 향상을 위한 제언 도출에 유용한 정보를 제공할 것으로 전망하며, 본 논문에서 제시한 분석 모델의 평가 결과 개선과 품질 향상 시스템 개발을 향후 과제로 남긴다.

References

[1] 국립중앙과학관 https://terms.naver.com/entry.naver?docId=3386304&cid=58370&categoryId=58370

[2] SAMSUNG SDS, 김서연, “알기쉬운 딥러닝-딥러닝과 빅데이터란 무엇인가?” https://www.samsungsds.com/kr/insights/1232588_4627.html

[3] 시사상식사전 https://terms.naver.com/entry.naver?docId=3484358&cid=43667&categoryId=43667

[4] etnews, 김군규, “[기고] 딥러닝 품질검사가 필요한 이유” https://www.etnews.com/20221102000051

[5] 박홍진, “딥러닝 기반의 스마트공장 운영 활동과 효과 관계분석”, 한국팜질경영학회, 2021

[6] 생화학백과 https://terms.naver.com/entry.naver?docId=5141772&cid=60266&categoryId=60266

[7] 지질학백과 https://terms.naver.com/entry.naver?docId=5960805&cid=61234&categoryId=61234