

A blue-toned background featuring a dense circuit board pattern. On the left side, there is a large, stylized 'AI' logo enclosed in a square frame, which is part of a larger integrated circuit chip design.

SK HYNIX X SCV

반도체 Sensor Data를 활용한 반도체 공정 품질 지수 가상 계측 모델 개발

SK Hynix 이주학 TL
이호진 김진수 안승아 임유경

CONTENTS

SEMICONDUCTOR

3

반도체 정의

반도체의 8대 공정

DATASET

11

Feature

Target

PREPROCESSING

12

전처리 과정

MODELING

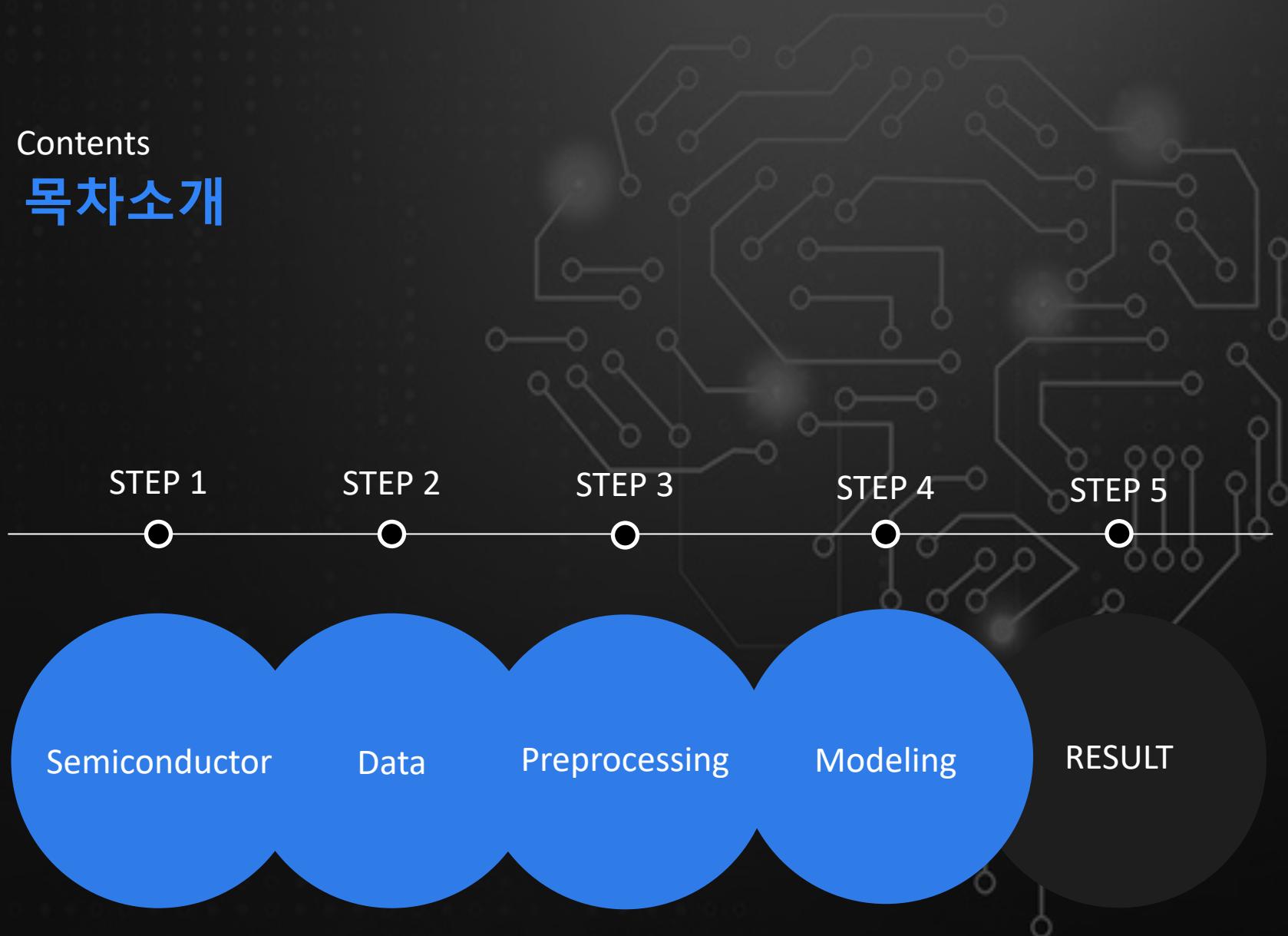
18

AUTO ML

Hyper Parameter Tuning

Contents

목차소개



01. Semiconductor

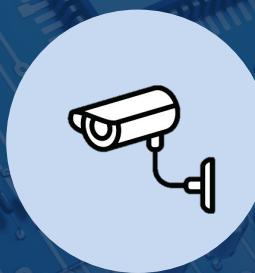
Semiconductor(반도체)

도체와 절연체의 역할 수행 (절연체의 전류 OFF 기능 + 도체의 전류 ON 기능)
→ 데이터 전환, 저장, 연산, 제어 하는 기능으로 사용

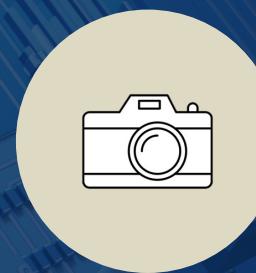
반도체 활용 영역



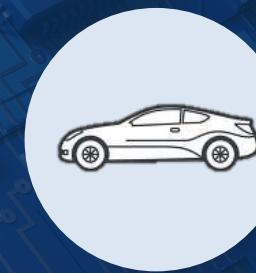
Mobile phone



CCTV



Camera



Automotive

01. Semiconductor

Q. 자율주행차 속 반도체에 문제가 생긴다면?

→ 인간의 목숨과 직결된 중요한 문제

따라서, **반도체 품질**이 매우 중요

→ 지금은 완벽한 품질 파악이 어려운 실정

Q. 품질 불량이란?

: Probe test는 통과했지만 고객에게 제공 시 불량이 발생하는 경우



01. Semiconductor

반도체 8대 공정

1. 웨이퍼 제조 공정	2. 산화 공정	3. 포토공정	4. 식각(Etch) 공정
5. 증착 & 이온주입 공정	6. 금속배선공정	7. EDS 공정	8. 패키징 공정

■ Inline 공정(FAB)
■ Test 및 포장 공정

01. Semiconductor

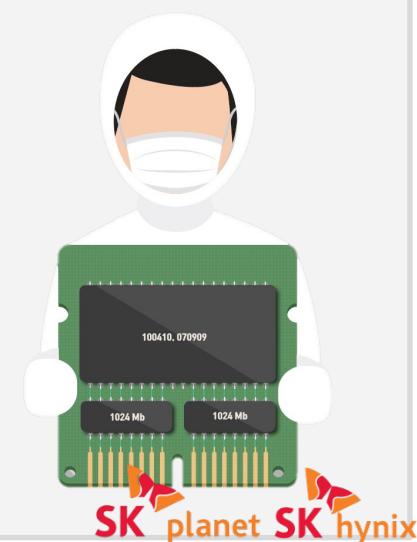
계측 : 제품이 생산되는 과정 중 제품이 잘 만들어지고 있는지 품질을 측정

Q. 계측이 필요한 이유?

→ 품질 확인 및 수율 개선

가상 계측 : Sampling 안 된 제품들에 대해 물리적인 측정 대신 장비 센서 등의 데이터에 기반해 가상으로 품질을 예측

> 현재 Wafer의 계측은 약 1%만 진행



01. Semiconductor

1. 웨이퍼 제조	2. 산화 공정	3. 포토공정	4. 식각공정	5. 증착 & 이온주입 공정	6. 금속배선공정
			<p>노광 김광액 산화막 웨이퍼</p> <p>식각 김광액 산화막 웨이퍼</p>	<p>웨이퍼 임플란트 산화막</p> <p>박막 박막</p>	

Q. 왜 Wafer의 계측은 약 1%만 진행될까?

Inline 공정은 한 Wafer당 약 700개의 공정 진행 * 수십 번의 계측 진행

따라서, 한 Wafer 계측을 위해 수천번의 계측 필요

모든 계측 진행 시

> 1) 생산성 저하 2) TAT 증가 3) 품질 악화

따라서, 현재 계측이 1%만 진행 중

→ 계측률을 높이기 위해 **가상계측(VM)** 꼭 필요

01. Semiconductor

목표

반도체 공정에 VM(Virtual Metrology)을 이용한 품질 지수 예측 모델 개발

제조 공정에서 발생하는 실제 데이터에 AI 기술을 활용한 가상 계측 모델
→ 실제 계측 장비와 유사한 예측 정확도를 나타냄



01. Semiconductor

Virtual Metrology for Thin Film process

현업 VM 측정 지표 RMSE : 약 7.5

→ 실제 산업에서 활용 되고 있는 검증된 VM을 바탕으로 결과 성능을 좀더 높이면 실제 산업에 즉각적인 개선 방안 적용 가능

→ 현업 모델을 고도화하기 위해 반도체 모델을 적용

02. DATASET

CVD(Chemical Vapor Deposition) 공정 중 다양한 **센서 데이터(Feature)** 생성



Amount of Gas



Distance



Power



Pressure



Temperature

Feature : Gas Power EPD HV 등 / Target : 박막의 두께

- > 박막의 두께와 굴절률은 반도체 품질과 직결
- > 박막이 미세하기 때문에 이를 계측하는 과정에 많은 시간과 자원이 소요
→ 전수 계측 어려움



03. Preprocessing

- 반도체 공정 이해를 통한 새로운 컬럼 생성(Step_id + param_alias)
→ pd.pivot_table 이용

PARAM_ALIAS (95개)	0	1	2
	EPD_para4	TIME_para7	GAS_para10

STEP_ID (7개)	0	1	2
	4	6	13



Step_param	04_EPD_para4	06_TIME_para7	13_GAS_para10
------------	--------------	---------------	---------------

→ 총 665개(95개 * 7개)의 Columns 생성



03. Preprocessing

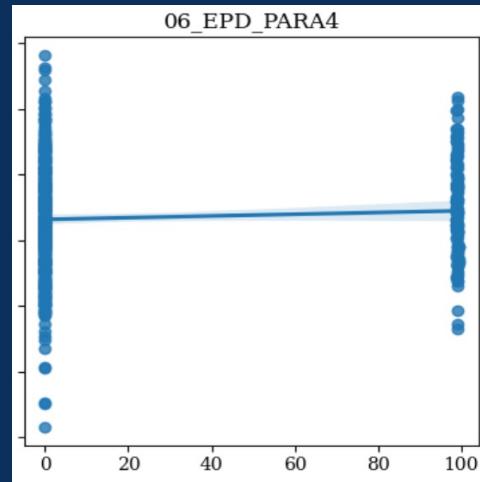
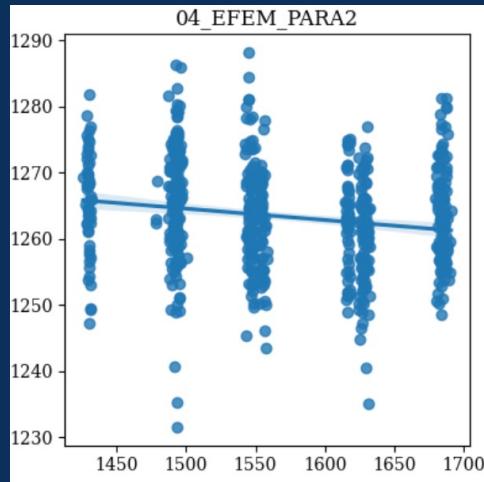
2. lst_stepsgap(공정별 소요시간)

총 7개의 step이 존재 [04, 06, 12, 13, 17, 18, 20]

> ex) [04~06] [06~12] [12~13]

> 각 step이 넘어갈 때 소요된 시간 계산 후 column 생성

3. 데이터 시각화를 통한 변수별 전처리



1) [A, B, C, D, E]로 구간화 후 One-hot encoding

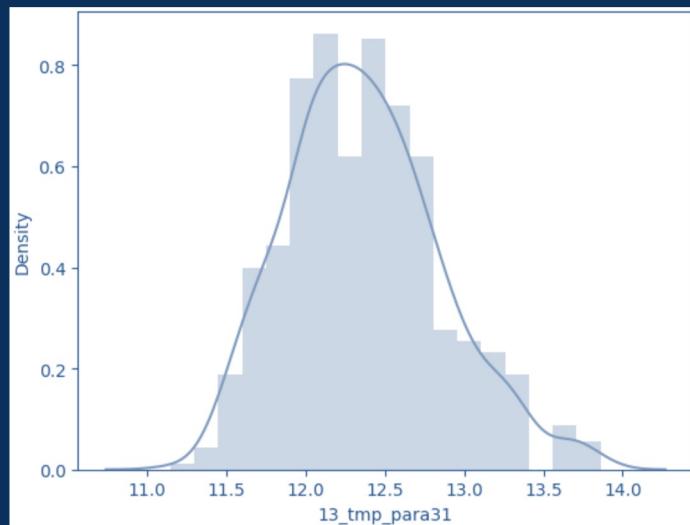
2) [A, B]로 구간화 후 One-hot encoding

03. Preprocessing

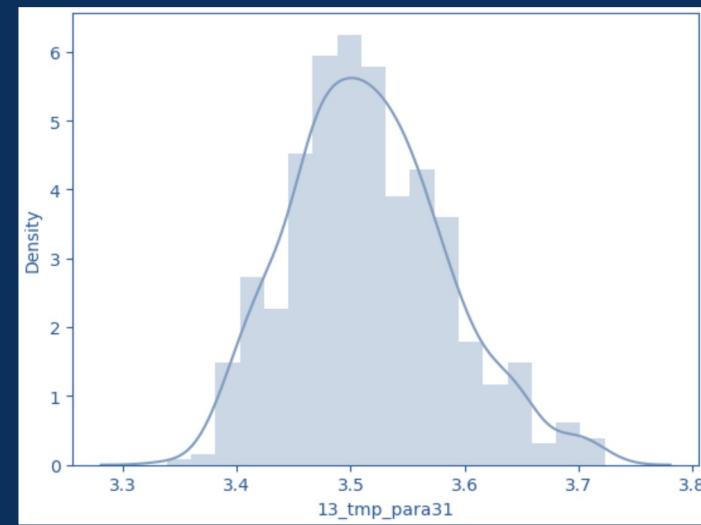
5. Para 이름 기준(temp, gas, efem 등)으로 평균, 표준편차 column 생성

gen_efem_para2_std	gen_efem_para25_std	gen_efem_para78_std
1.540604	1.487185	0.045480
1.313742	1.372572	0.045224

6. 전체 데이터의 왜도 높은 변수 -> 정규분포와 가깝게 하기 위한 log scaling 처리



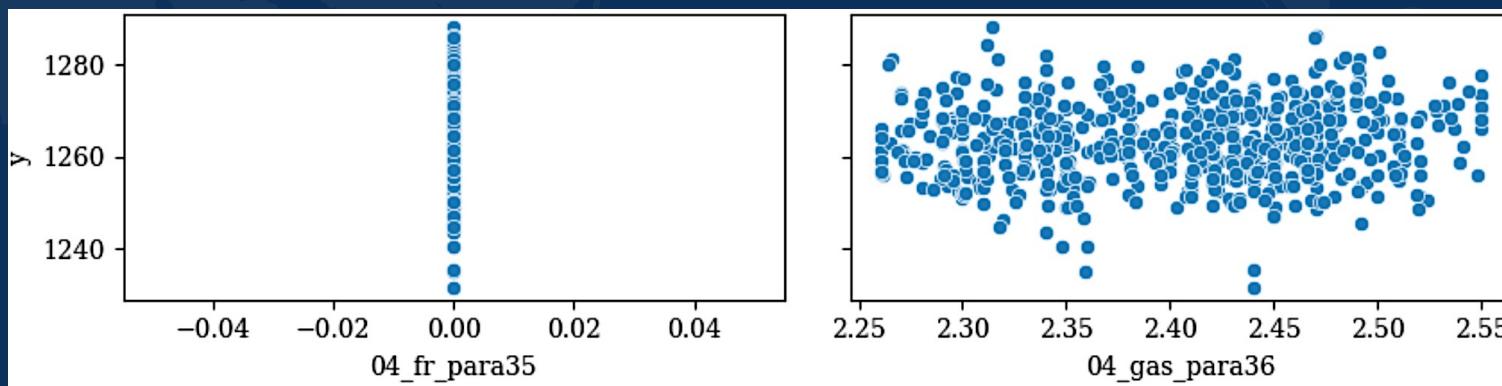
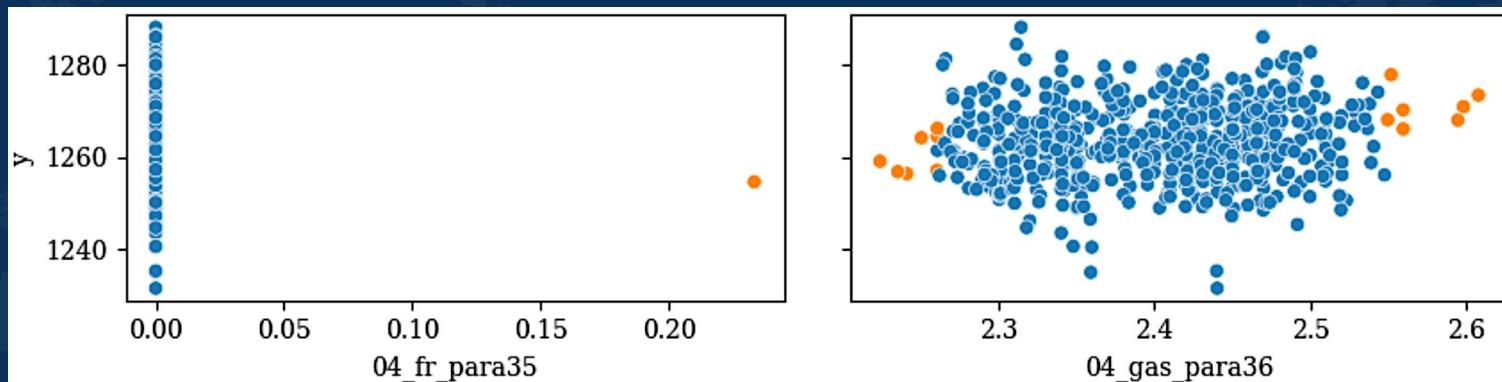
Skew : 0.4533



Skew : 0.3996

03. Preprocessing

7. Outlier 처리

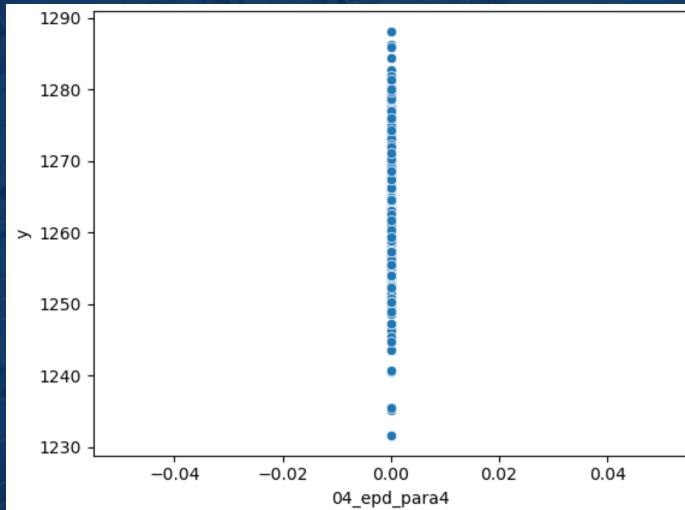


03. Preprocessing

8. 분산 처리

VarianceThreshold 이용한 분산 0 제거

> variance가 지정한 threshold 보다 작으면 drop하는 방법



20_temp_para54	0.000000e+00
gen_tmdiff_1718	0.000000e+00
gen_epd_para80_std	0.000000e+00
gen_he_para1_std	1.975385e-31
gen_he_para95_std	3.160616e-30

분산이 0인 컬럼 뿐 아니라 분산이 0과 가까운 컬럼 모두 제거

03. Preprocessing

9. 원-핫 인코딩

Module_name	Module_name_eq_EQ1	Module_name_eq_EQ10	Module_name_eq_EQ11
EQ1	1	0	0
EQ10	0	1	0
EQ11	0	0	1

10. 다중공선성 처리 및 PCA 적용

다중공선성 : 회귀분석에서 독립변수들 간에 강한 상관관계가 나타나는 문제

VIF(Variance Inflation Factor) 10이상인 변수 선정 후 제거 (276개)

VIF가 10 이상? > 독립변수 없이도 충분히 반응변수(y) 설명 가능

VIF Factor	Features
11+1.133280e	17_fr_para61
10+4.805839e	12_time_para64
10+4.011151e	20_time_para34

04. Modeling

1. 변수 선택



```
SelectKBest(score_func = mutual_info_regression, k = 250)
```

Columns 중 유의미한 변수 선정 필요

→ **SelectKBest** 활용

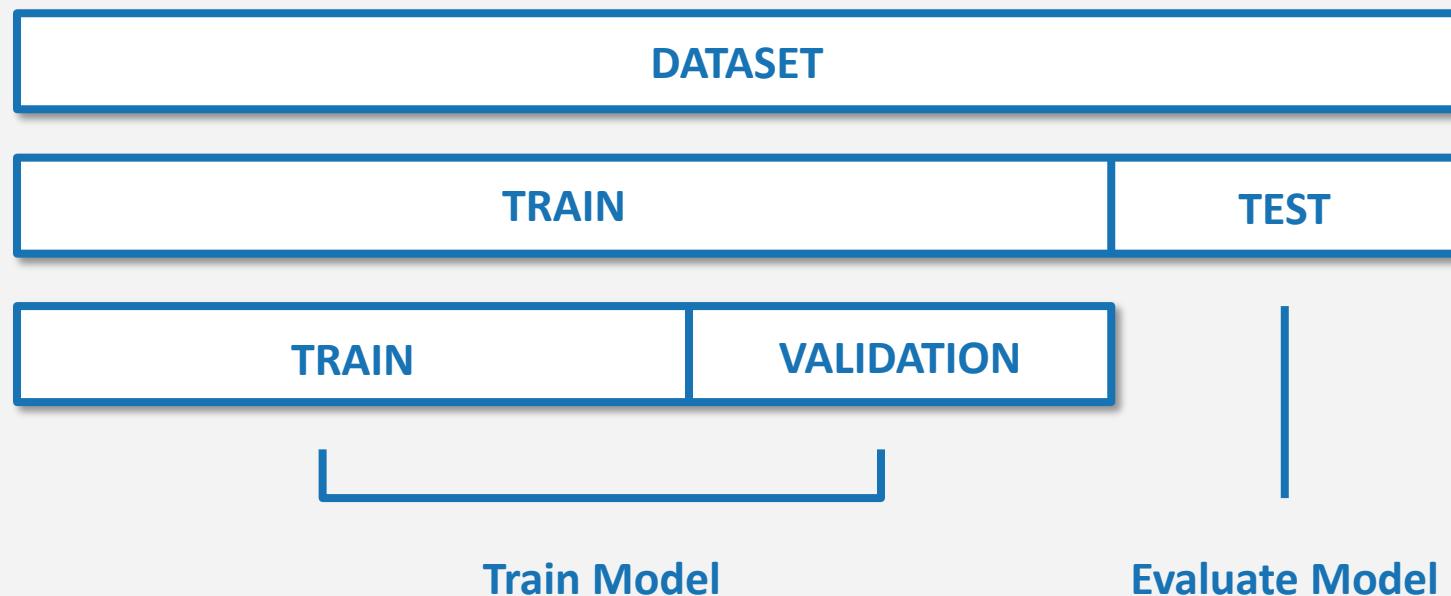
→ Target과 상관 관계가 높은 feature 250개 선택

`mutual_info_regression` : Feature 간의 독립성을 고려해 측정

04. Modeling

2. 데이터셋 분리

Train : Valid : Test = 0.64 : 0.16 : 0.2



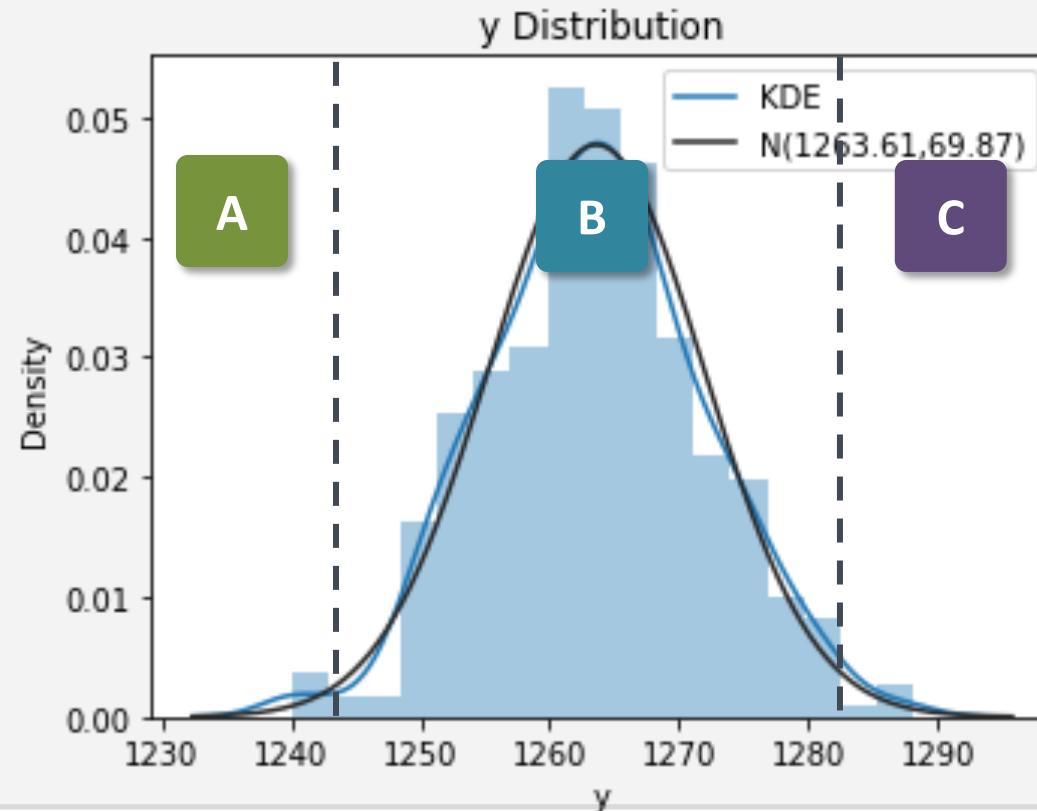
04. Modeling

3. 랜덤 오버 샘플링

최종 Train data에 대해 target 값의 빈도가 극단적으로 낮은 구간 → [A, B, C]로 구분

Random Oversampling 진행

→ Minor 구간의 instance를 임의 복제해 A, B, C 의 균형성을 맞춤



04. Modeling

4. Auto ML



전처리 및 변수 선택 후 PYCARET Auto ML 진행



데이터 준비



모델 생성 및 비교



하이퍼파라미터 튜닝



학습된 모델 분석

자동 모델 분석 진행

04. Modeling

5. HPT(Hyper Parameter Tuning)

Auto ML 진행 후

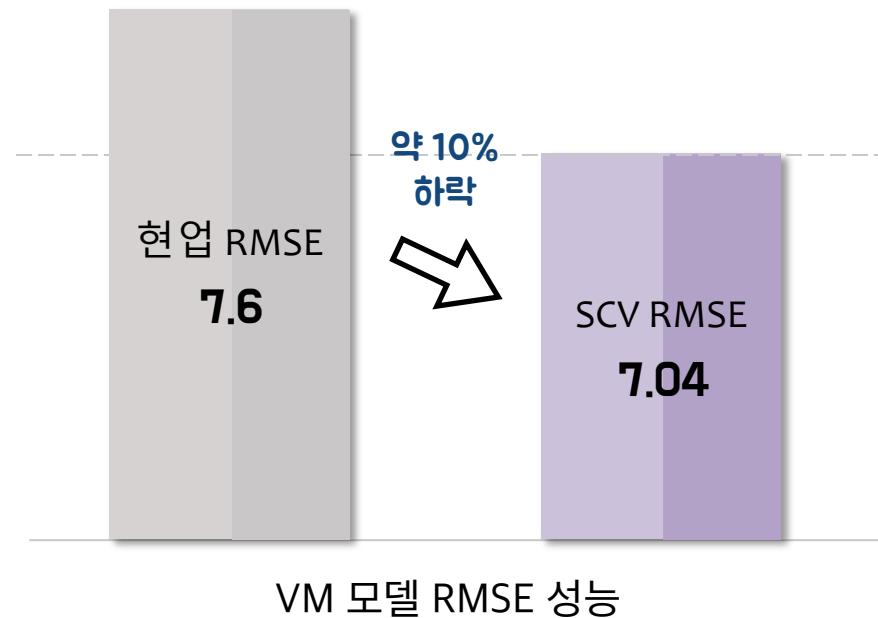
RMSE 성능 Top 3 모델 → HPT 진행

Model		MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
rf	Random Forest Regressor	2.1215	20.0631	4.4193	0.9425	0.0035	0.0017	1.0990
et	Extra Trees Regressor	2.1216	20.6112	4.4942	0.9411	0.0036	0.0017	0.8650
lightgbm	Light Gradient Boosting Machine	2.1899	20.8833	4.5122	0.9404	0.0036	0.0017	0.3070
catboost	CatBoost Regressor	2.1985	20.8584	4.5221	0.9404	0.0036	0.0017	24.6620

RMSE SCORE	AUTO ML (Ensemble)	Random Forest	Extra Trees	LightGBM
7.04				

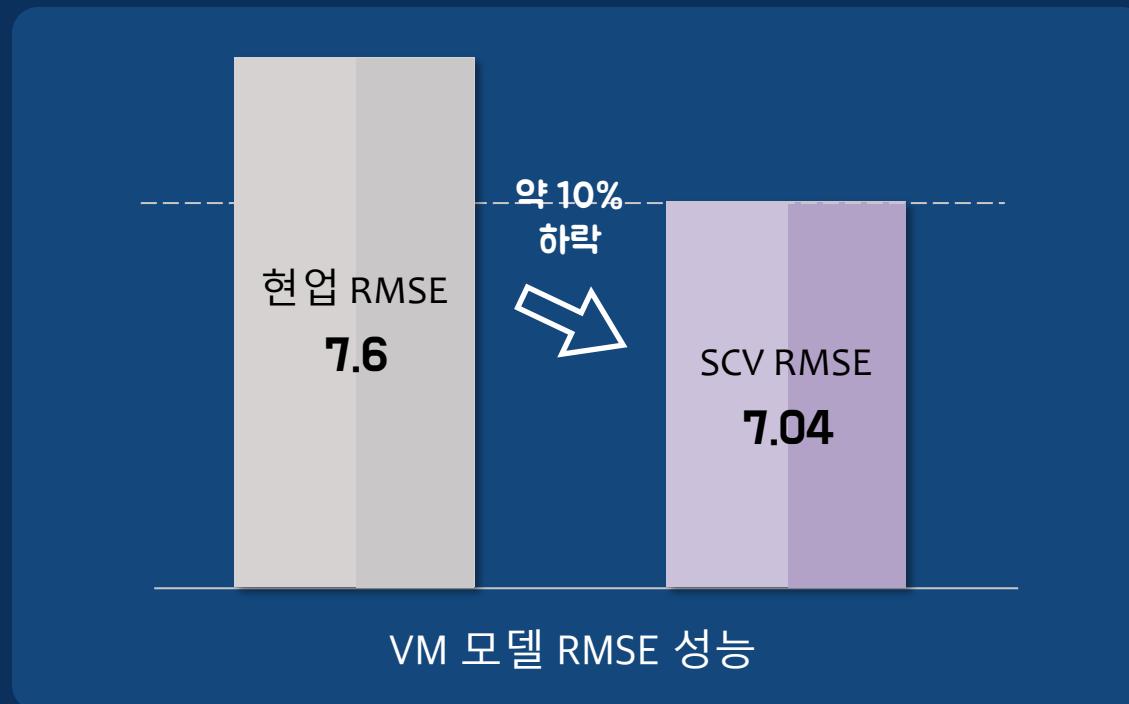
RESULT

RMSE 하락에 따른 기대효과



- > 공정 산포 감소 및 수율 향상에 도움
- > VM 계측장비 투자비용 줄일 수 있음

05. RESULT



- > 공정 산포 감소 및 수율 향상에 도움
- > VM 계측장비 투자비용 줄일 수 있음

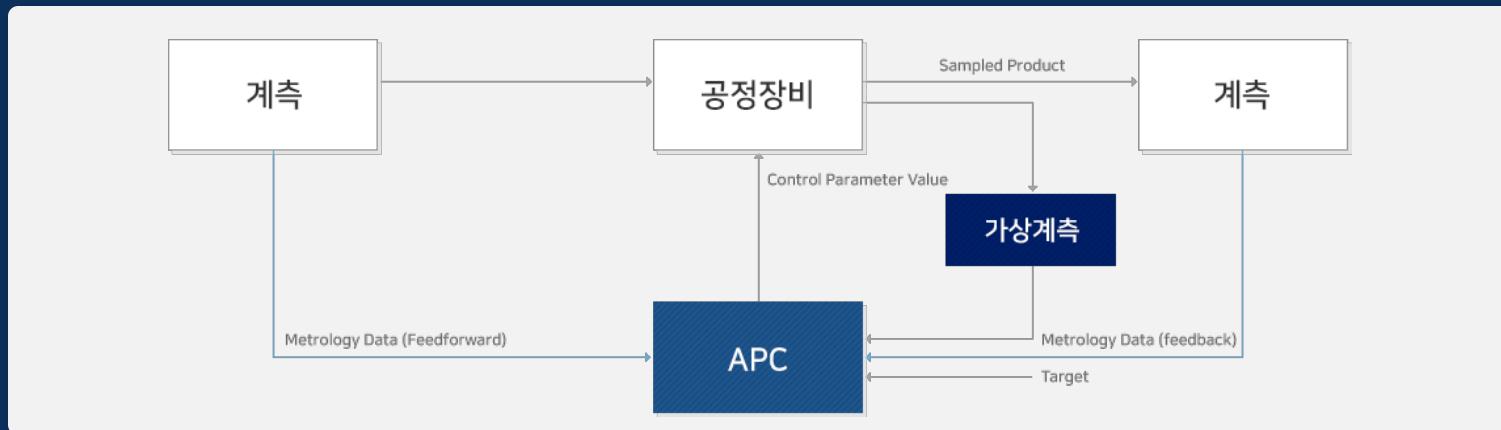
05. RESULT

VM 기대효과

1) 제조사가 100% 전수 검사하는 것과 같은 효과

2) VM 결과 값을 APC(Advanced Process Control)와 연동

> 장비의 최적 공정 조건을 찾아주는 솔루션(APC)을 통해 수율 향상



05. RESULT

VM 기대효과

3) 인라인 전수 데이터 분석을 통한 수율 ramp-up 기간 단축

* ramp-up : 생산 능력의 증가

4) 계측 장비 투자 비용 절감

5) 객관적인 wafer 품질 Grading 을 통한 품질불량 방지