



Introduction to Business Analytics: Case Study

강필성

고려대학교 산업경영공학부

pilsung_kang@korea.ac.kr

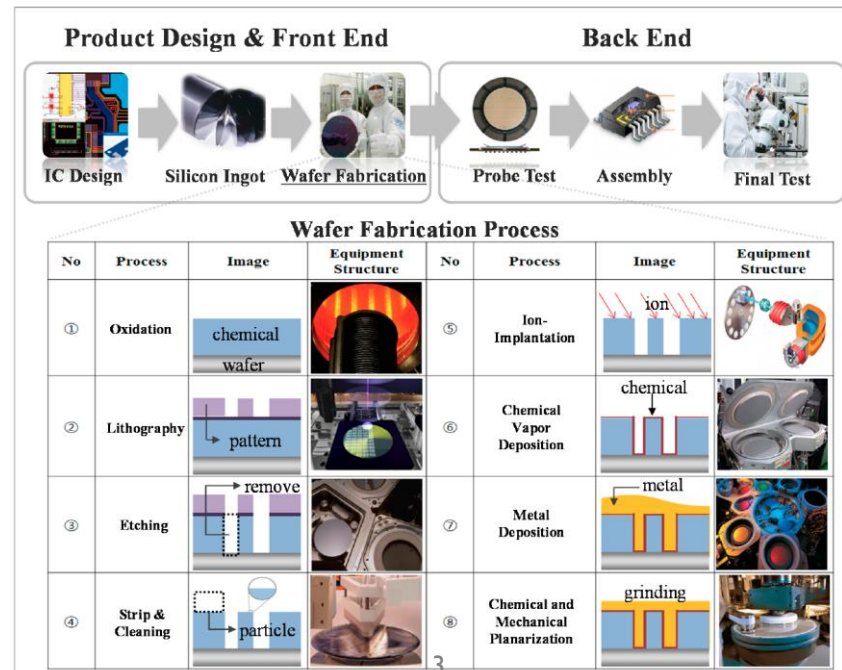
AGENDA

- 01 빅데이터 분석 개요 및 주요 개념
- 02 데이터 과학 프로젝트 절차
- 03 기계 학습 방법론
- 04 **제조업 활용 사례: 가상 계측 모델 개발**

Virtual Metrology based on FDC Data

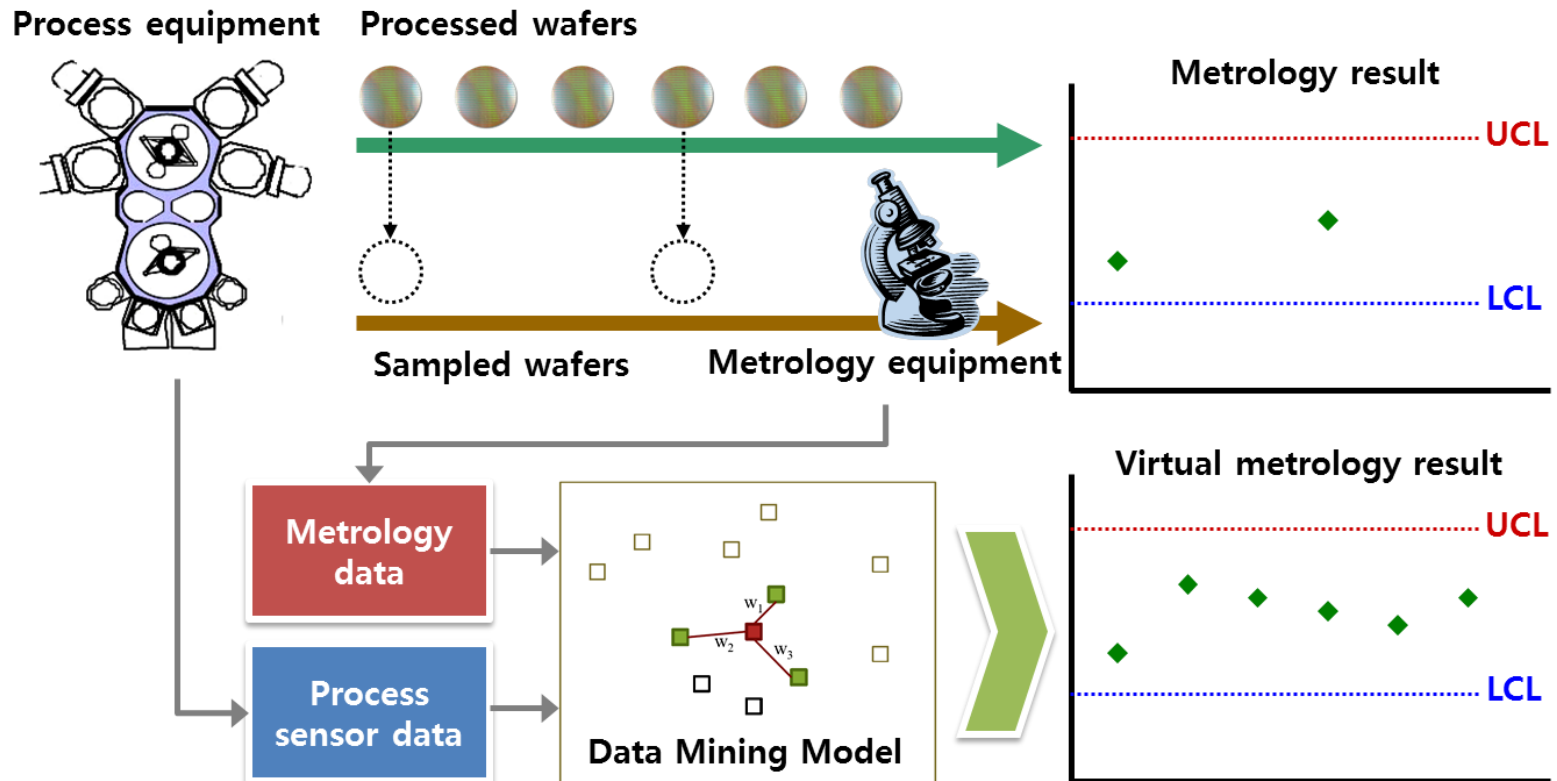
• 문제 인식

- ✓ 반도체 공정은 100개 이상의 세부 공정으로 이루어져 있으며, FAB-IN에서 FAB-OUT 까지 평균 45일 가까운 시간이 소요 (자동차 72시간, 철강 48시간 이내)
- ✓ 주요 공정 이후에 품질관리를 위해 계측을 수행
 - 샘플링 기반의 검사이므로 Type I/II 오류 발생
 - 계측 검사 기간에 소요되는 시간만큼의 Delay 발생



Virtual Metrology based on FDC Data

- 시도 I: FDC 데이터를 사용한 가상계측 모델 개발



Kang et. al. (2009) A virtual metrology system for semiconductor manufacturing, *Expert Systems with Applications* 36(10): 12554-12561.

Virtual Metrology based on FDC Data

- 데이터 탐색 및 전처리

- ✓ 대상 공정: Etching

- FDC Summary variables

Etching process	Machine-chamber	Equipment sensors	Available variables
Etching1	8	48	192
Etching2	18	56	224

- Target variables

System	Target	UCL	Mean	LCL
VM1	Target1	1.052	1.000	0.948
	Target2	1.060	1.000	0.970
VM2	Target3	1.056	1.000	0.972
	Target1	1.200	1.000	0.800

Virtual Metrology based on FDC Data

- 데이터 탐색 및 전처리

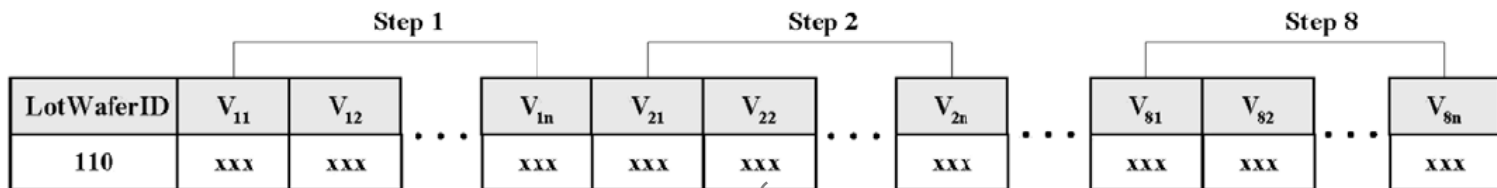
- ✓ FDC data transformation

- 참고: 최근에는 시간 정보를 반영하여 CNN 또는 RNN 구조를 사용하기도 함

(a) Before Reconstruction(Matrix Form)

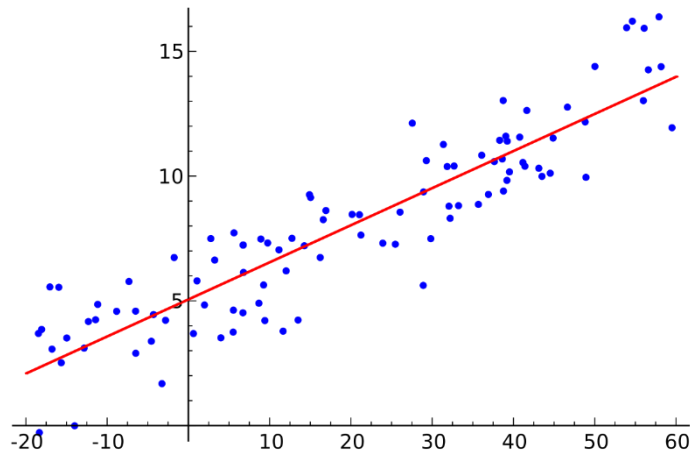
LotID	WaferID	StepID	V_1	V_2	...	V_n
1	10	1	xxx	xxx	...	xxx
1	10	2	xxx	xxx	...	xxx
1	10	3	xxx	xxx	...	xxx
1	10	4	xxx	xxx	...	xxx
1	10	5	xxx	xxx	...	xxx
1	10	6	xxx	xxx	...	xxx
1	10	7	xxx	xxx	...	xxx
1	10	8	xxx	xxx	...	xxx

(b) After Reconstruction(Vector Form)

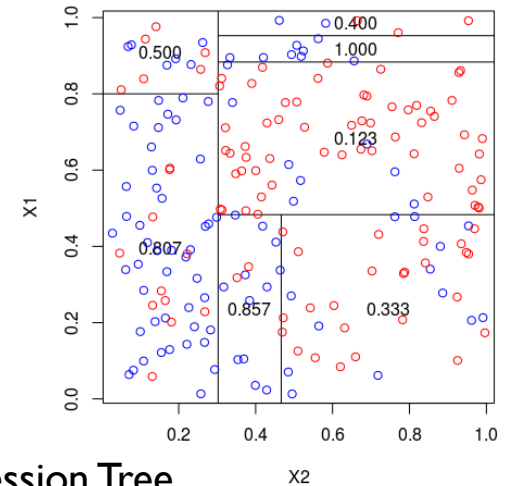
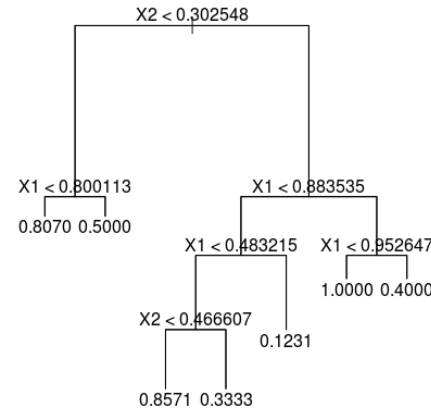


Virtual Metrology based on FDC Data

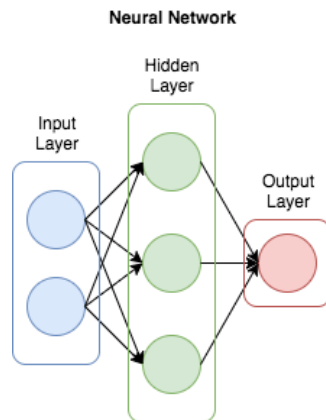
- Prediction Models



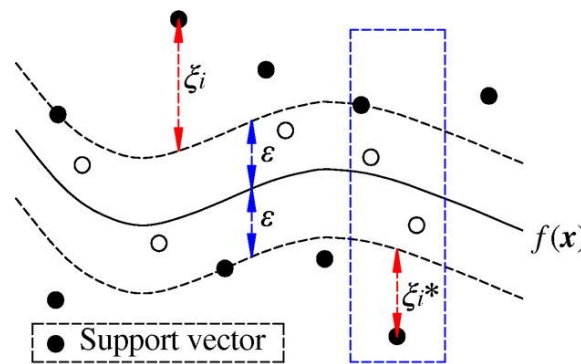
Multiple Linear Regression (MLR)



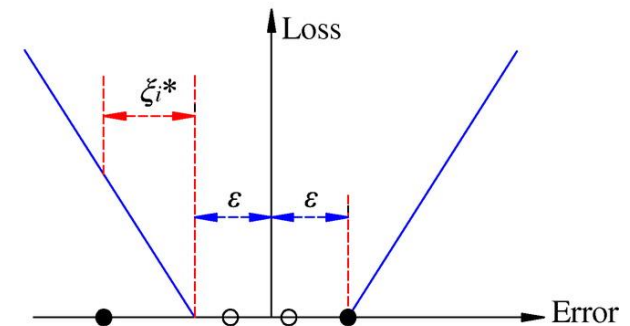
Regression Tree



Neural Network



(a)



(b)

Support Vector Regression (SVR)

Virtual Metrology based on FDC Data

- 차원 축소: Dimensionality Reduction

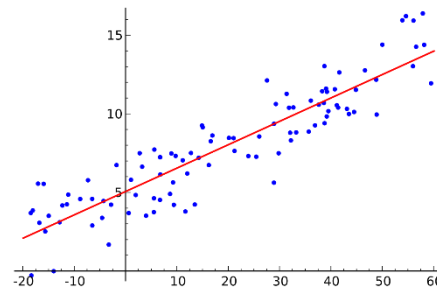
- ✓ VMI Model: 118 wafers with 1,536 variables

- ✓ VM2 Model: 241 wafers with 1,792 variables

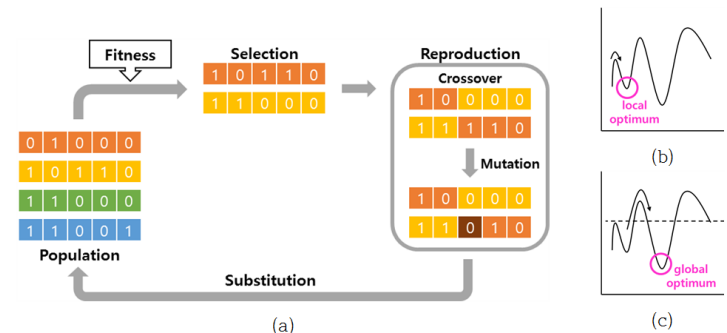
- ✓ 가용한 웨이퍼 수에 비해 FDC Sensor 변수 수가 너무 많음

- 차원 축소 방법론

- ✓ 변수 선택

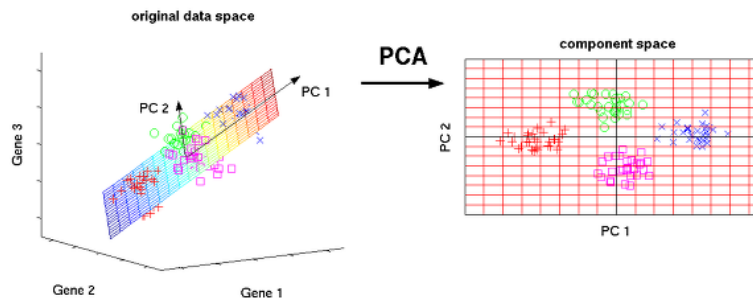


Multiple Linear Regression (MLR)

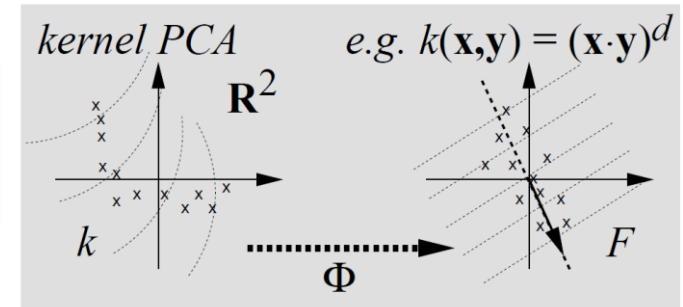


GA-SVR

- ✓ 변수 추출



Principal Component Analysis (PCA)



Kernel Principal Component Analysis (KPCA)

Virtual Metrology based on FDC Data

- 차원 축소 결과

✓ 실질적으로 VM 모델을 구축하는데 사용된 FDC 변수의 수는 최소 8개, 최대 239개

Table 6

The number of input variables derived from each dimensionality reduction method for each target (A(B + C), A: total number of variables, B: the number of variables from equipment sensors, C: the number of variables from metrology data).

Method	VM1			VM2
	Target1	Target2	Target3	Target1
LR-A	34 (24 + 10)	22 (12 + 10)	34 (24 + 10)	21 (20 + 1)
LR-S	10 (9 + 1)	8 (7 + 1)	12 (11 + 1)	21 (20 + 1)
GASVR-A	106 (96 + 10)	239 (229 + 10)	219 (209 + 10)	154 (153 + 1)
GASVR-S	72 (69 + 3)	48 (47 + 1)	90 (86 + 4)	122 (122 + 0)
PCA-A-70	30 (20 + 10)	30 (20 + 10)	30 (20 + 10)	23 (22 + 1)
PCA-S-70	20	20	20	22
KPCA-A-70	30 (20 + 10)	30 (20 + 10)	30 (20 + 10)	24 (23 + 1)
KPCA-S-70	21	21	21	23

Virtual Metrology based on FDC Data

• VM 예측 오차

✓ 당시 산학과제 의뢰 기업에서 제시했던 허용 오차의 50%를 넘지 않음

VM1 Target1	Regression model				
Dim. reduction	LR	k-NN	RT	NN	SVR
LR-A	0.0297	0.0333	0.0345	0.0342	0.0254
LR-S	0.0145	0.0365	0.0375	0.0184	0.0156
GASVR-A	71.458	0.0464	0.0352	0.0365	0.0314
GASVR-S	28.827	0.0366	0.0332	0.0389	0.0243
PCA-A-70	0.0458	0.0468	0.0436	0.0436	0.0291
PCA-S-70	0.0669	0.0637	0.0927	0.0457	0.0565
KPCA-A-70	0.0460	0.0476	0.0425	0.0328	0.0291
KPCA-S-70	0.0649	0.0661	0.1064	0.0587	0.0565

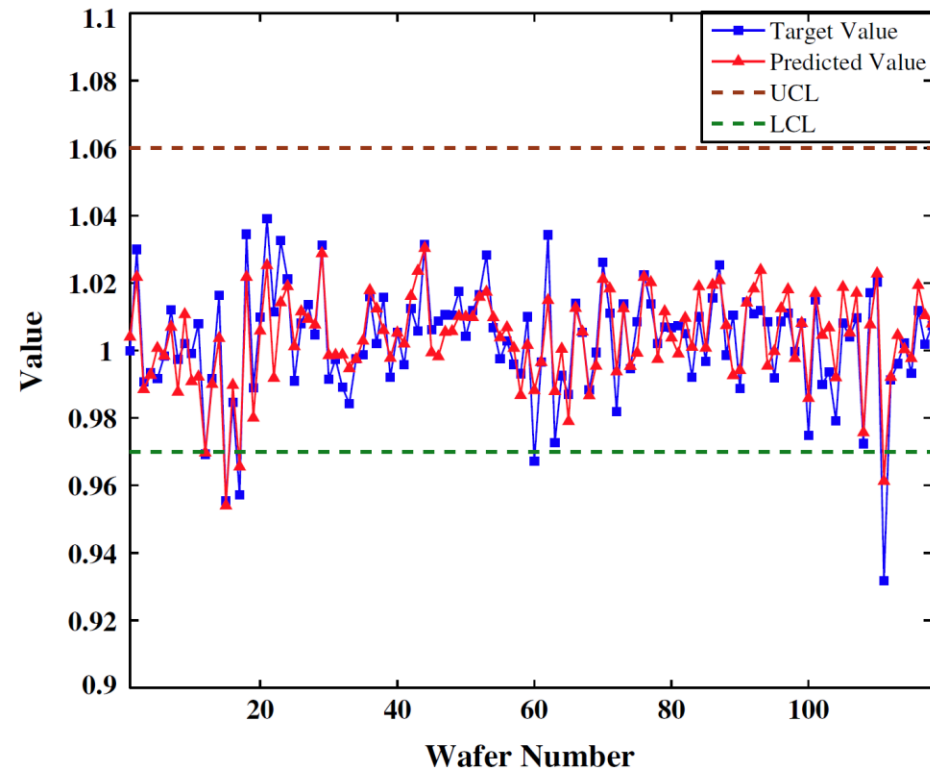
VM1 Target2	Regression model				
Dim. reduction	LR	k-NN	RT	NN	SVR
LR-A	0.0121	0.0114	0.0174	0.0132	0.0093
LR-S	0.0053	0.0111	0.0123	0.0148	0.0063
GASVR-A	0.0351	0.0201	0.0139	0.0508	0.0150
GASVR-S	5.5767	0.0175	0.0113	0.0133	0.0089
PCA-A-70	0.0153	0.0146	0.0145	0.0180	0.0088
PCA-S-70	0.0289	0.0263	0.0455	0.0248	0.0208
KPCA-A-70	0.0153	0.0145	0.0141	0.0134	0.0087
KPCA-S-70	0.0268	0.0244	0.0369	0.0226	0.0208

VM1 Target3	Regression model				
Dim. reduction	LR	k-NN	RT	NN	SVR
LR-A	0.0186	0.0259	0.0364	0.0262	0.0200
LR-S	0.0116	0.0260	0.0346	0.0172	0.0143
GASVR-A	0.0859	0.0362	0.0556	0.0502	0.0293
GASVR-S	5.4033	0.0401	0.0575	0.0307	0.0271
PCA-A-70	0.0283	0.0335	0.0493	0.0283	0.0245
PCA-S-70	0.0353	0.0406	0.0542	0.0371	0.0307
KPCA-A-70	0.0282	0.0336	0.0496	0.0359	0.0243
KPCA-S-70	0.0350	0.0402	0.0662	0.0338	0.0309

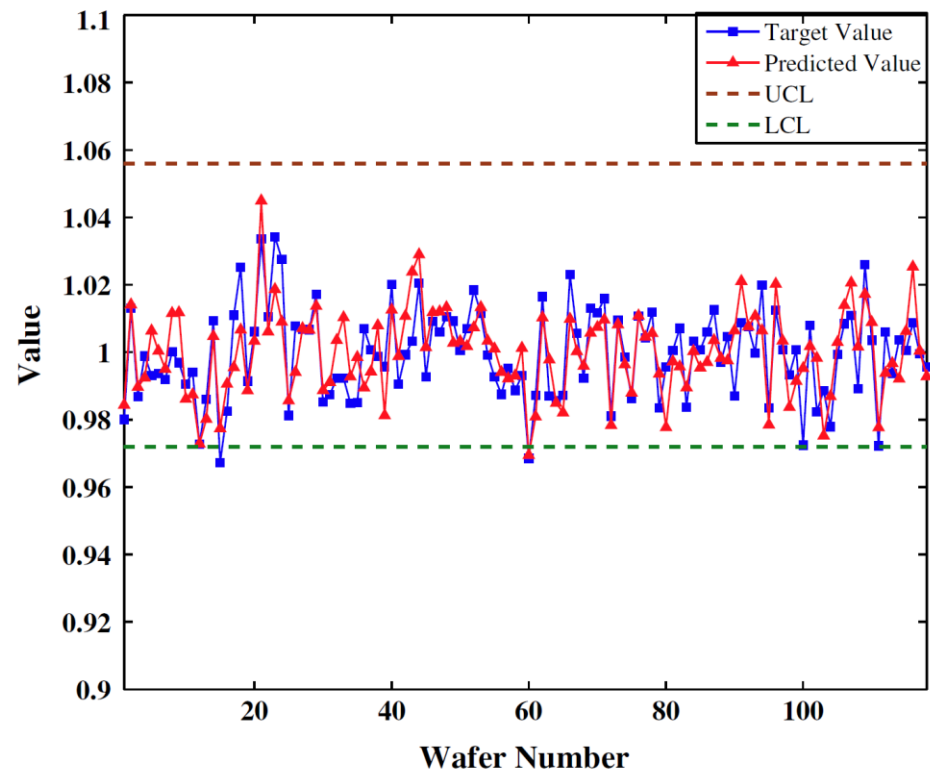
VM2 Target1	Regression model				
Dim. reduction	LR	k-NN	RT	NN	SVR
LR-A	0.0218	0.0503	0.0496	0.0243	0.0254
LR-S	0.0197	0.0388	0.0532	0.0247	0.0252
GASVR-A	5.9268	0.0321	0.0541	0.0346	0.0277
GASVR-S	0.6150	0.0358	0.0544	0.0388	0.0300
PCA-A-70	0.0375	0.0671	0.0614	0.0369	0.0353
PCA-S-70	0.0375	0.0673	0.0572	0.0374	0.0367
KPCA-A-70	0.0378	0.0727	0.0691	0.0418	0.0353
KPCA-S-70	0.0392	0.0761	0.0596	0.0395	0.0367

Virtual Metrology based on FDC Data

- VM 예측 예시



(a) Target2 of VM1.



(b) Target3 of VM1.

Virtual Metrology with Prediction Reliability

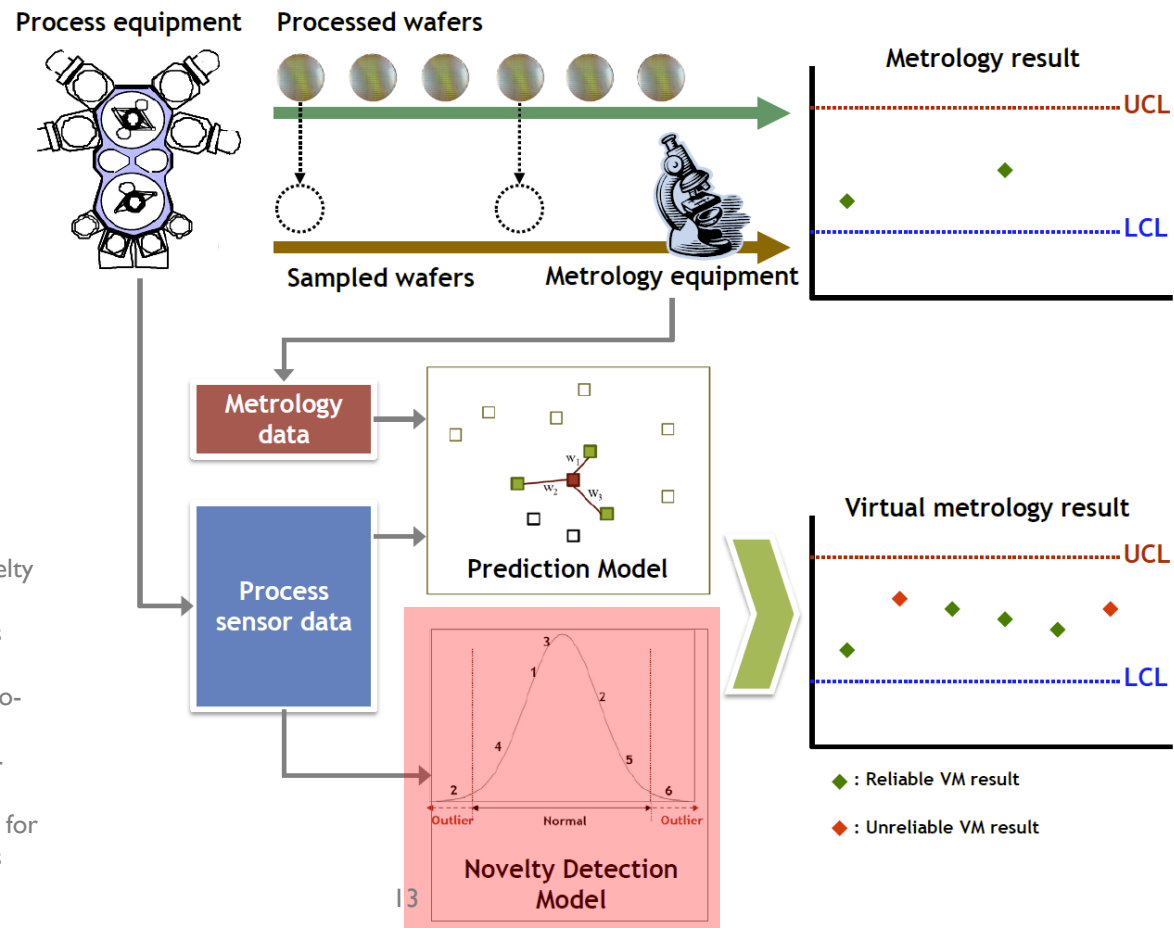
- 시도 I의 문제점

- ✓ 실제 공정에서 PoC를 수행한 결과 False Alarm이 자주 발생
- ✓ 원인: 학습에서 사용했던 FDC 데이터와 예측 모델 적용시 발생하는 FDC 데이터 간의 분포 차이 존재

Virtual Metrology with Prediction Reliability

• 시도 2

- ✓ 대안: 예측 모형의 신뢰도를 함께 제공하자
- ✓ 높은 신뢰도를 갖는 예측 결과는 그대로 사용하고, 아닐 경우 엔지니어가 개입



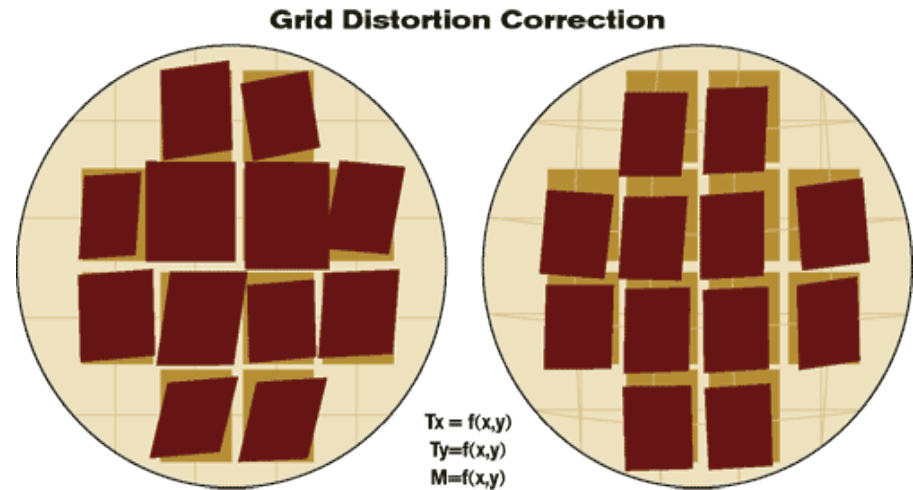
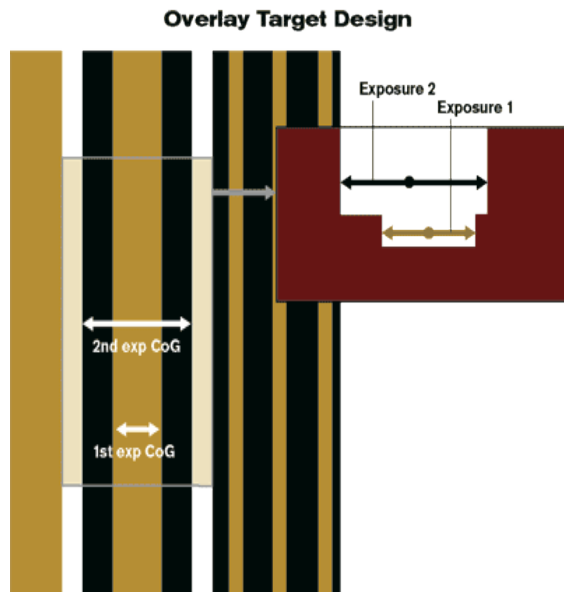
- Kim et al. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications* 39(4): 4075-4083.
- Kang et al. (2011). Virtual metrology for run-to-run control in semiconductor manufacturing. *Expert Systems with Applications* 38(3): 2508-2522.
- Kang et al. (2009). A virtual metrology system for semiconductor manufacturing. *Expert Systems with Applications* 36(10): 12554-12561.

Virtual Metrology based on FDC Data

• 데이터 탐색 및 전처리

✓ 대상 공정: Photolithography

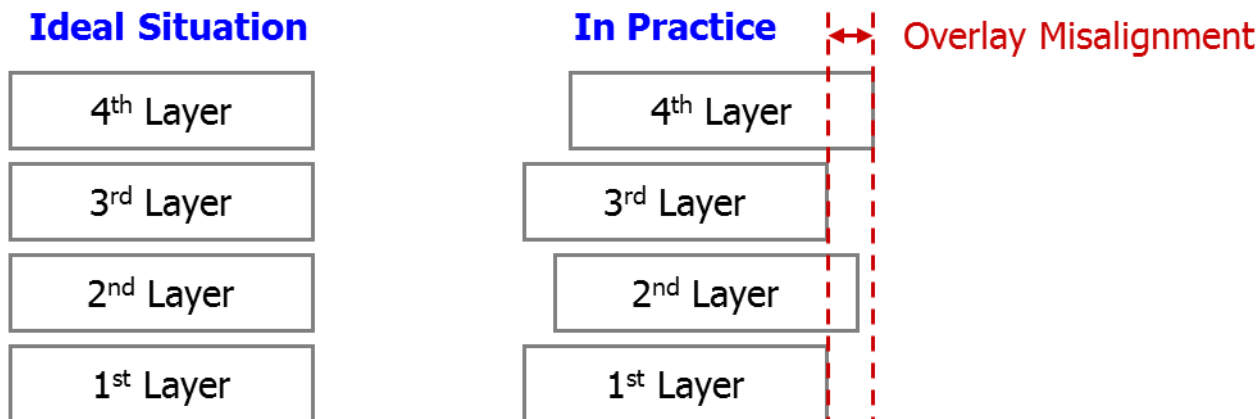
- 웨이퍼(Wafer)에 빛을 쏘아 반도체 회로를 새기는 공정
- 하나의 웨이퍼 위에 여러 번에 걸쳐 회로를 새김
- 확대, 축소, 변형, 회전 등의 어긋남에 의해 최종 수율(yield)이 목표값에 비해 하락



Virtual Metrology based on FDC Data

- 데이터 탐색 및 전처리

✓ 계측 정보: Overlay misalignment (Target 변수: 8개)



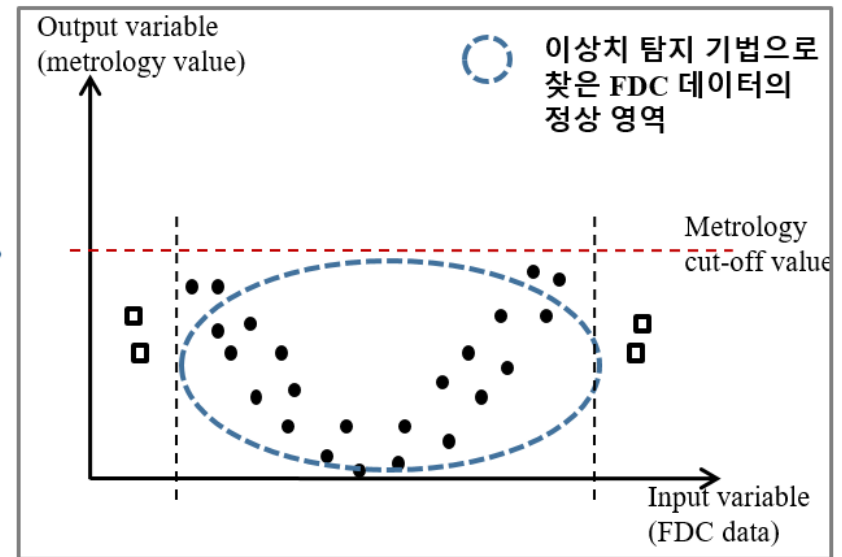
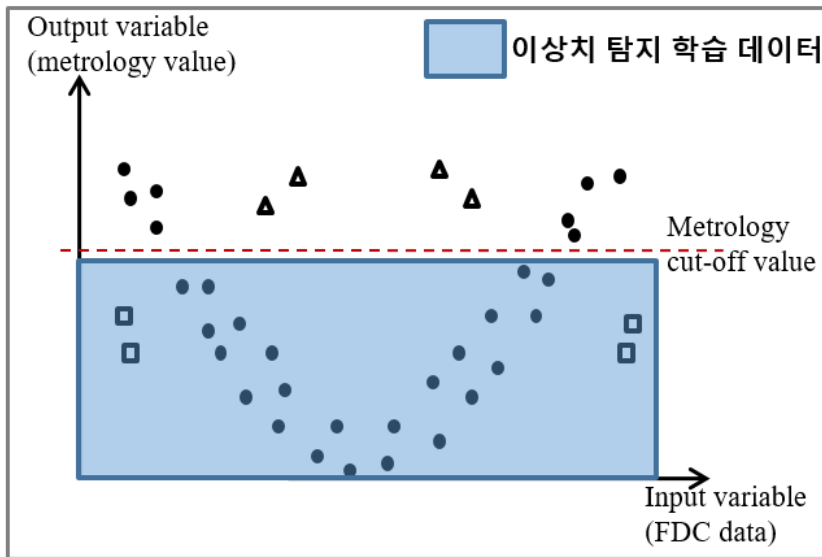
✓ FDC Data: 두 대의 장비로부터 수집된 4개월치 데이터, PM 기준으로 총 8가지 기간 존재, 133개의 FDC 입력 변수와 4개의 출력 변수 사용

	P1	P2	P3	P4	P5	P6	P7	P8
EQ1	230	172	137	167	452	818	138	195
EQ2	226	180	136	170	450	816	138	195

■ 참고: PM과 관계 없는 모델을 만들고 싶었으나 PM 전후로 존재하는 급격한 shift 처리 실패

Virtual Metrology based on FDC Data

- 이상치 탐지를 통한 학습 데이터 정제 및 신뢰도 추정
 - ✓ 출력 변수의 정상/비정상 여부는 사전 기준이 존재
 - ✓ 입력변수의 정상/비정상 영역은 정해져 있지 않음
 - ✓ 출력변수가 정상인 입력변수 데이터에 대해 데이터의 90%를 설명하고 나머지 10%를 이상치로 판단하는 이상치 탐지 모델 학습

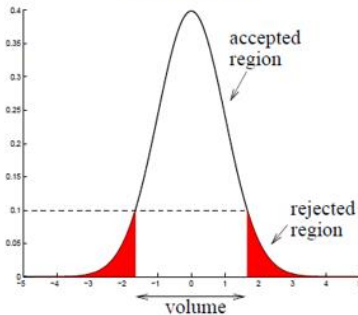


Virtual Metrology based on FDC Data

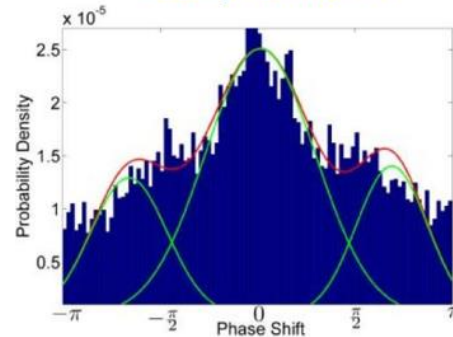
• 이상치 탐지

✓ 총 일곱 가지의 이상치 탐지 알고리즘을 사용하여 노이즈 데이터 판별

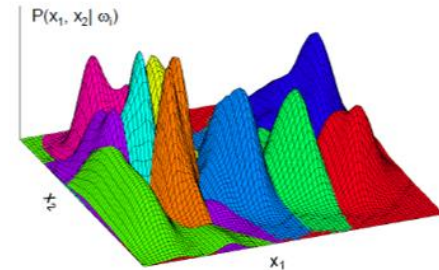
(1) Gaussian density estimation



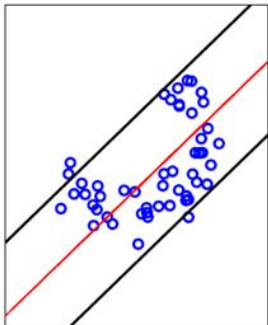
(2) Mixture of Gaussian density estimation



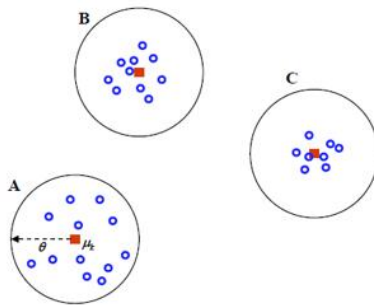
(3) Parzen window density estimation



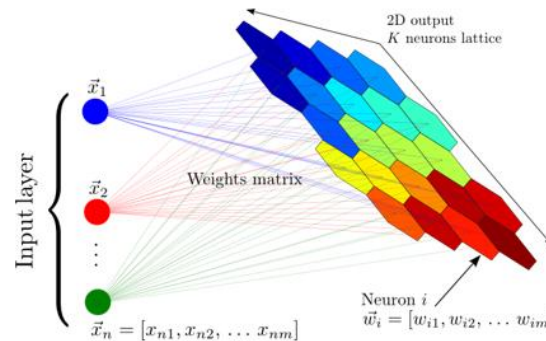
(4) PCA



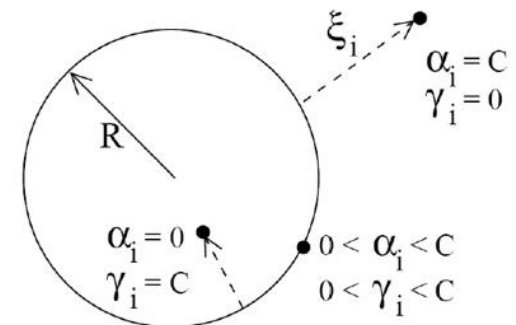
(5) K-Means Clustering



(6) Self-organizing Map



(7) SVDD



Virtual Metrology based on FDC Data

- 이상치 탐지

✓ 노이즈 제거 방법

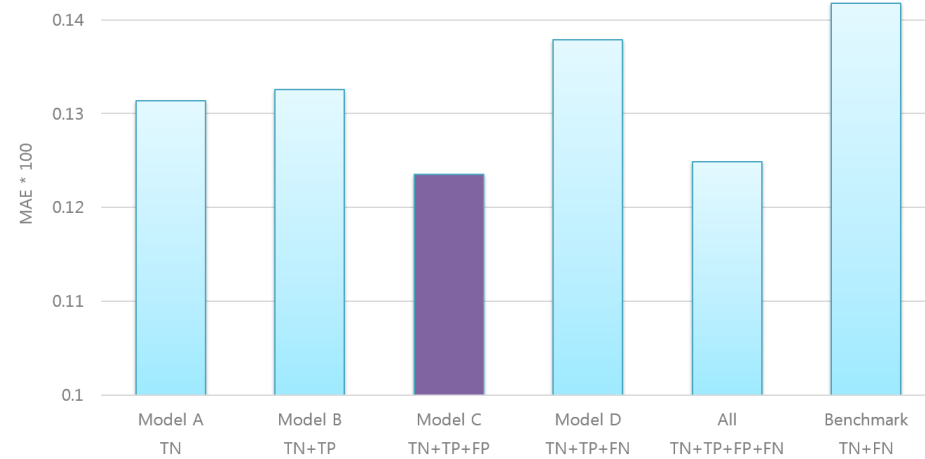
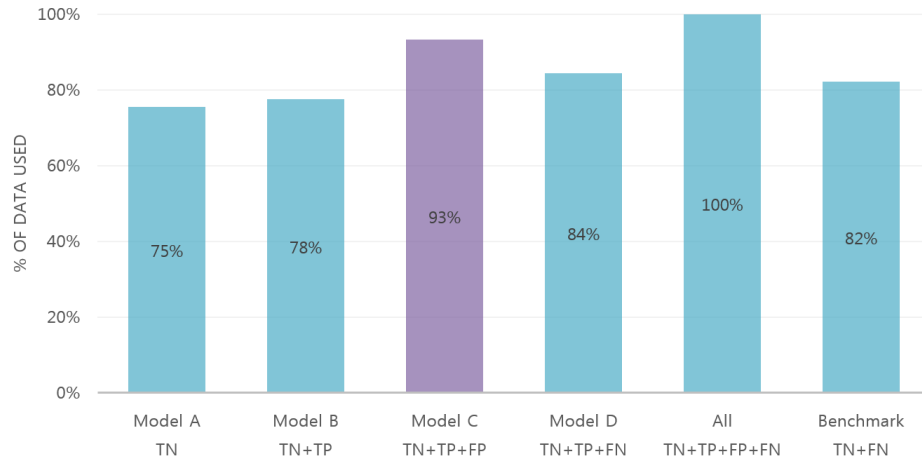
입력변수 이상치 탐지 결과

		정상	이상치(비정상)
실제 계측결과	정상	True Negative	False Positive (내부 노이즈)
	비정상	False Negative (외부 노이즈)	True Positive

분류	사용 데이터	설명
Model A	TN	정상 데이터만 사용
Model B	TN+TP	노이즈를 모두 제거
Model C	TN+TP+FP	전체 데이터에서 외부 노이즈를 제거
Model D	TN+TP+FN	전체 데이터에서 내부 노이즈를 제거
All	TN+TP+FP+FN	모든 데이터 사용
Benchmark	TN+FN	일반적인 노이즈 제거 방법론

Virtual Metrology based on FDC Data

- 학습 데이터에서 노이즈 제거 비율(좌) 및 가상계측 성능(우)



✓ 전체 데이터에서 외부 노이즈만 제거한 경우 전체의 **93%** 데이터를 이용하여 가장 낮은 오류(MAE)값을 보임

Virtual Metrology based on FDC Data

- 가상 계측 신뢰도 추정 예시

- ✓ 검정색 네모: 실제 계측 값, 파란색 다이아몬드: 신뢰도가 높게 부여된 웨이퍼, 붉은색 동그라미: 신뢰도가 낮게 부여된 웨이퍼

