



공정관리를 위한 빅데이터 활용현황

김성범

데이터마이닝 및 품질공학 연구실
고려대학교 산업경영공학과

- 소개
- 빅데이터 모델링 개요
- 공정 모니터링 사례
- 빅데이터 도입에 대한 의견



김성범

- 고려대학교 산업경영학과 교수 (2009 ~)
- 미국 텍사스주립대학교 교수 (2005 ~ 2009)
- 미국 조지아공과대학교 산업시스템공학 박사 (2005)

연구분야



- 데이터마이닝 / 머신러닝 알고리즘
- 장비/부품 수요예측, 고장예측
- 사물인터넷 기반 스마트공장 디자인
- 데이터기반 제조공정 품질/수율 향상

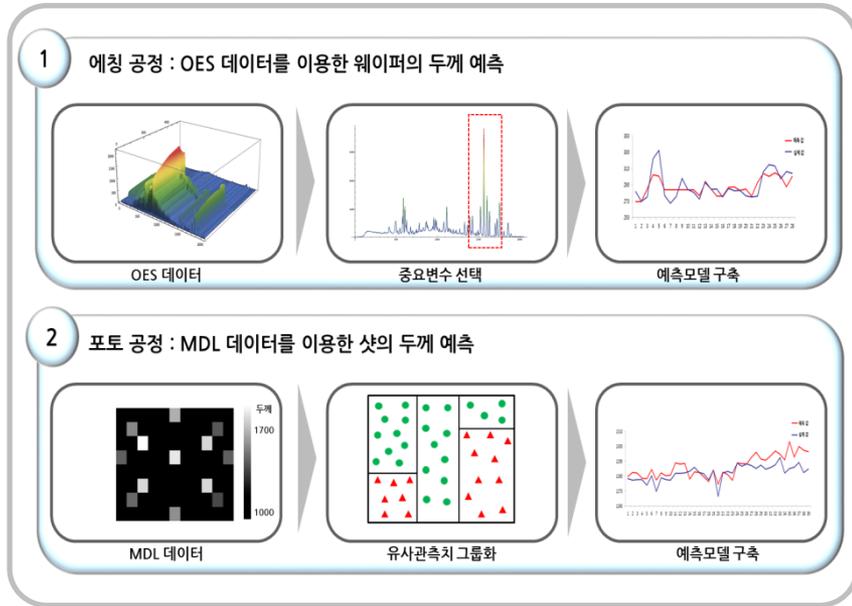
활동 및 수상



- 100 여개 국제 우수 논문 게재, 150 여회 국내외 학회 및 포럼 발표
- 한국데이터마이닝학회, 대한산업공학학회, 품질경영학회 이사
- 미국경영공학회 (INFORMS) 데이터마이닝 분과 회장
- 데이터마이닝 교육 – 삼성전자, LG 디스플레이, SK 하이닉스, 현대중공업
- 고려대학교 2014년도 크림슨 교수 선정 (우수신진교수)
- 고려대학교 강의우수상 (석탑강의상) 12회 수상

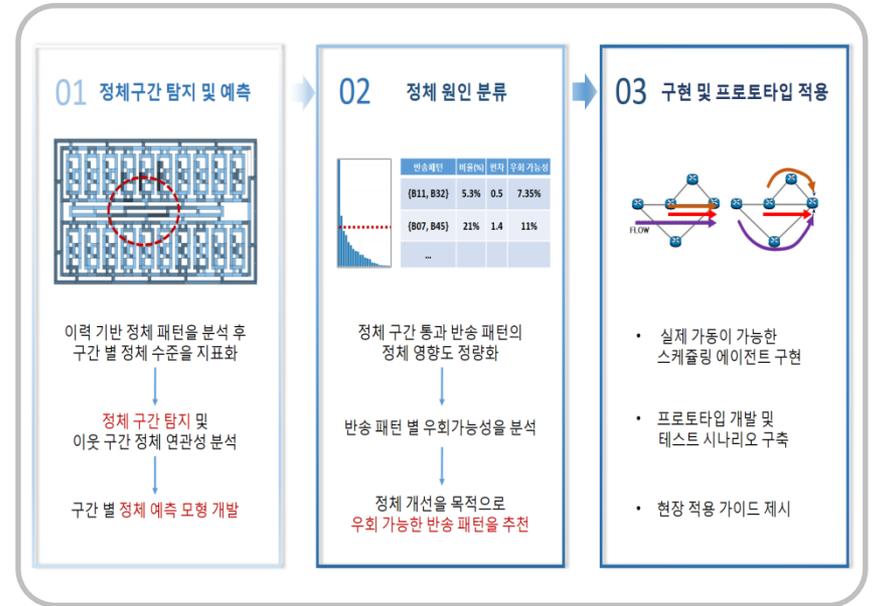
빅데이터 분석 실제 산학 프로젝트

삼성전자 In Fab의 Engineering Big Data를 활용한 공정 건강도 관리 방법론 개발



- 진보된 공정의 건강도 관리 방법을 도출
- 생산라인 별 품질분석 및 머신러닝 기반 가상계측 모델
- 공정에 영향을 미치는 주요 인자 추출

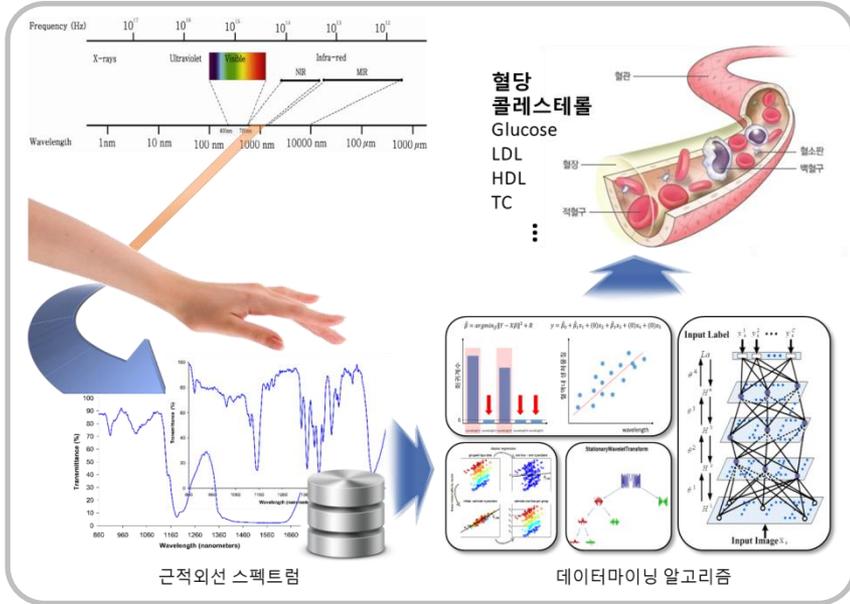
삼성전자 반도체 물류시스템 정체 예측 기법 개발



- 물류시스템 정체 지표 정의
- 정체 예측 프레임 구축
- 정체 원인 분류 및 우회 반송 패턴 추천

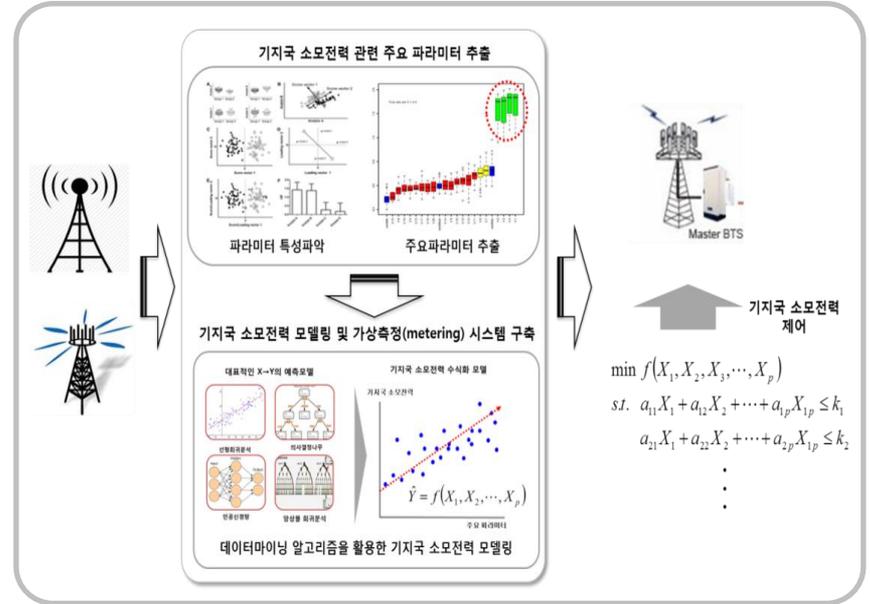
빅데이터 분석 실제 산학 프로젝트

삼성전자 생체정보 분석 및 예측 알고리즘 개발



- 근적외선 분광기 스펙트럼을 통한 혈중성분/혈당 분석
- 혈액 내 특정 생체물질 정량화 및 예측 알고리즘 개발

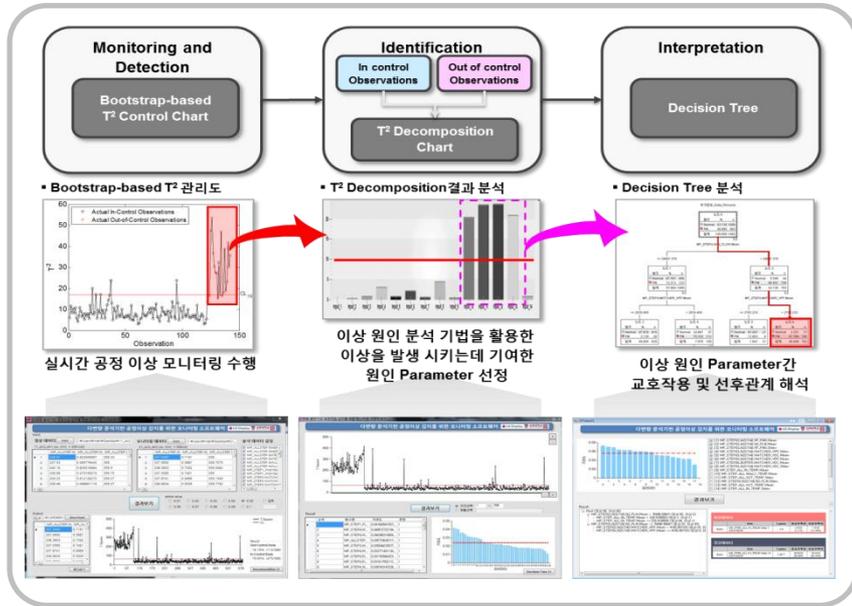
삼성전자 기지국 소모전력 예측 모델링



- 130개 기지국의 전력량을 예측하는 모델 구축
- 현업에서 적용 가능한 수준의 정확도 확보

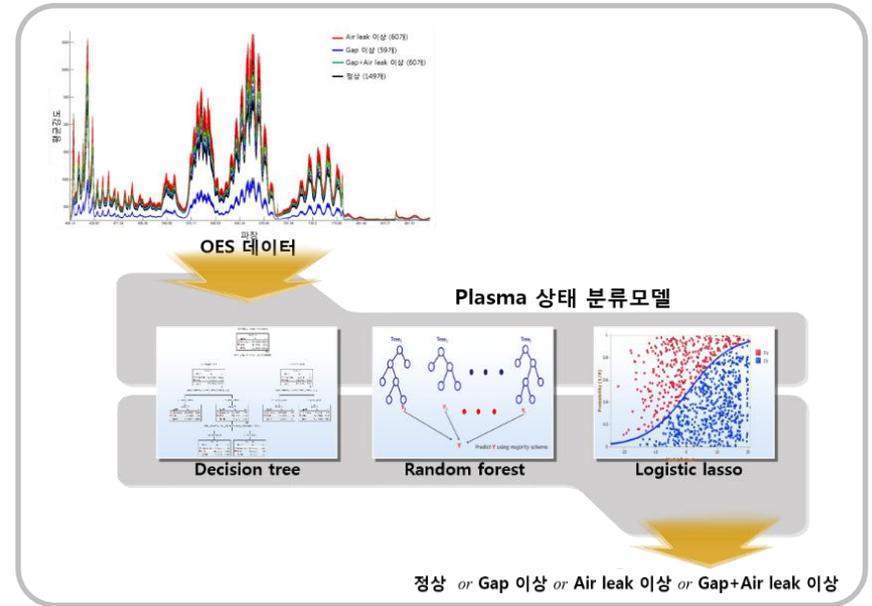
빅데이터 분석 실제 산학 프로젝트

LG 디스플레이 다변량 분석 기반 공정 이상 감지를 위한 모니터링 모델 개발 및 양산적용 효과산출 연구



- 변수간 상관관계를 고려하는 다변량 모니터링 모델 구축
- 체계적인 공정 이상원인 분석 프로세스를 제안
- 양산 적용을 통해 신속한 설비이상의 원인파악/개선

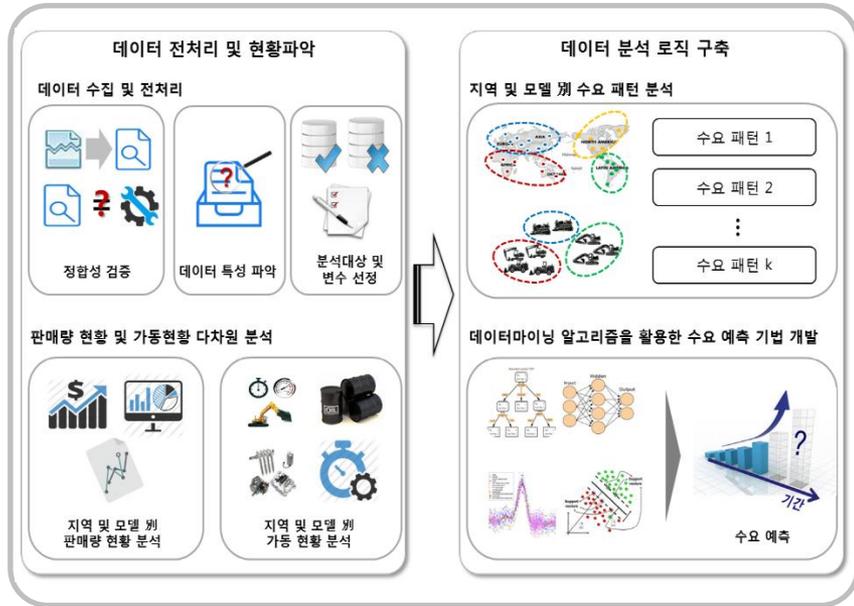
LG 디스플레이 Plasma 상태 진단을 위한 분류모델 구축



- OES 데이터를 이용하여 Plasma 진단 분류모델 구축
- 중요 변수를 제공함으로써 Plasma 상태 변동에 대한 중요 원인 파악

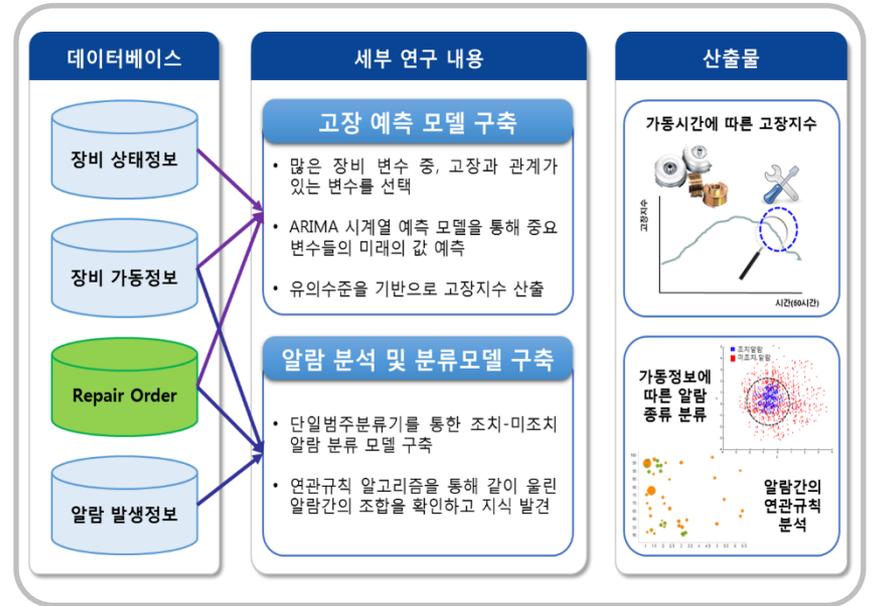
빅데이터 분석 실제 산학 프로젝트

현대중공업 빅데이터를 활용한 건설장비 수요예측



- 지역, 장비종류 별 가동패턴을 반영한 수요예측 모델 구축
- 유가, 환율 등 외부변수를 반영하여 예측모델 고도화

현대중공업 빅데이터를 기반 건설장비 품질분석기법 개발



- 건설장비 가동 정보와 실제 고장과의 관계 분석
- 가동에 따른 고장지수 산출 모델 구축
- 다양한 알람 중 실제 고장과 관련이 높은 중요 알람 선별

- 소개
- **빅데이터 모델링 개요**
- 공정 모니터링 사례
- 빅데이터 도입에 대한 의견

다변량 데이터

변수 관측치	X_1	...	X_i	...	X_p
N_1	x_{11}	...	x_{1i}	...	x_{1p}
...
N_i	x_{i1}	...	x_{ii}	...	x_{ip}
...
N_n	x_{n1}	...	x_{ni}	...	x_{np}

- 변수: 각 관측치의 특성치
- 단변량 데이터: 변수의 수 = 1
- 다변량 데이터: 변수의 수 ≥ 2

다변량 데이터 형태

ID	X_1	X_2	X_3	...	X_P
1					
2					
...					
N					

다변량

ID	V_1	V_2	V_3	...	V_P
1					
2					
...					
N					

시그널 (프로파일, 커브)

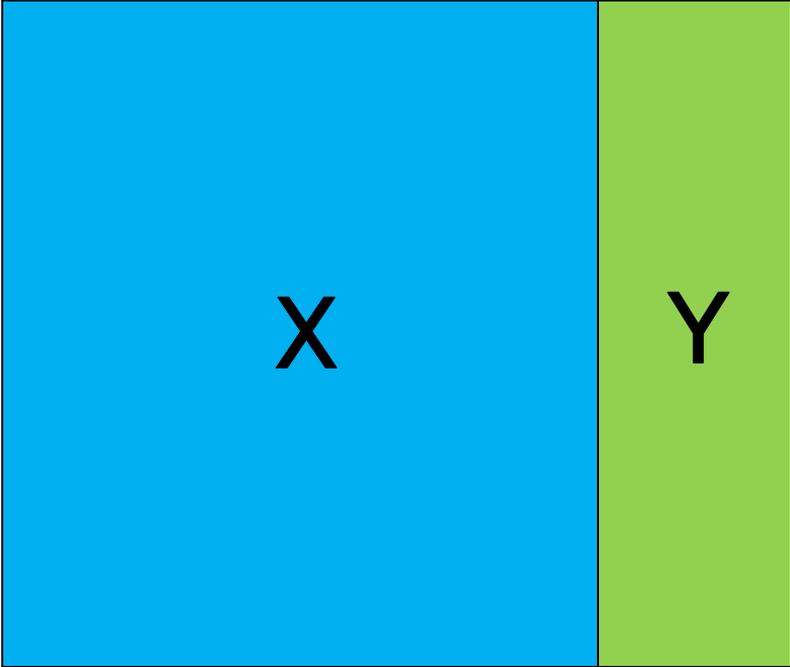
ID	x_1	x_2	x_3	...	x_P
1					
2					
...					
N					

네트워크

ID	V_1	V_2	V_3	...	V_P
1					
2					
...					
N					

이미지

데이터 구조



독립변수
예측변수
인풋변수

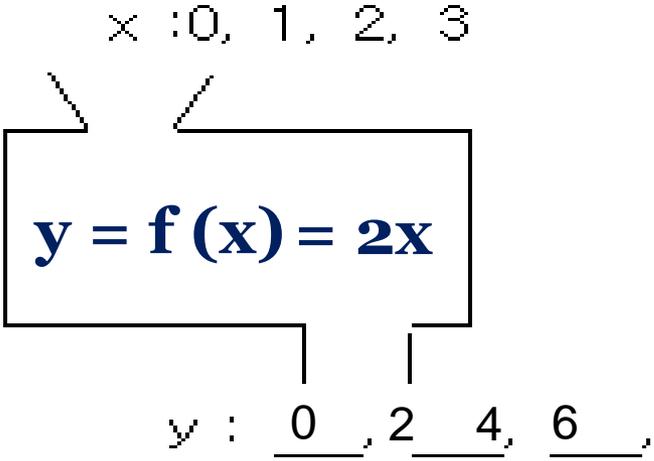
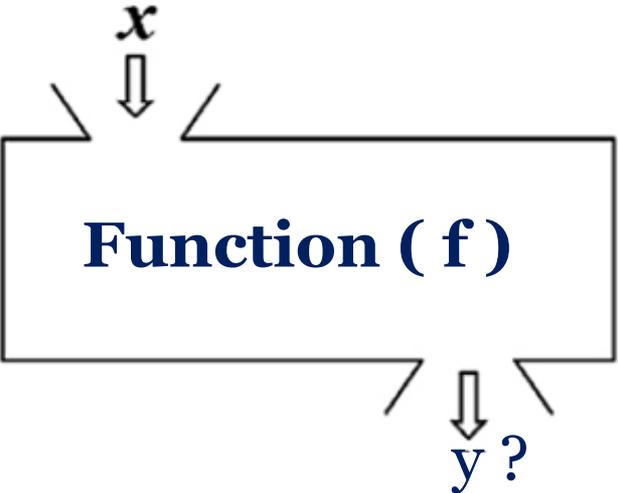
종속변수
반응변수
아웃풋변수

다변량 데이터 예제

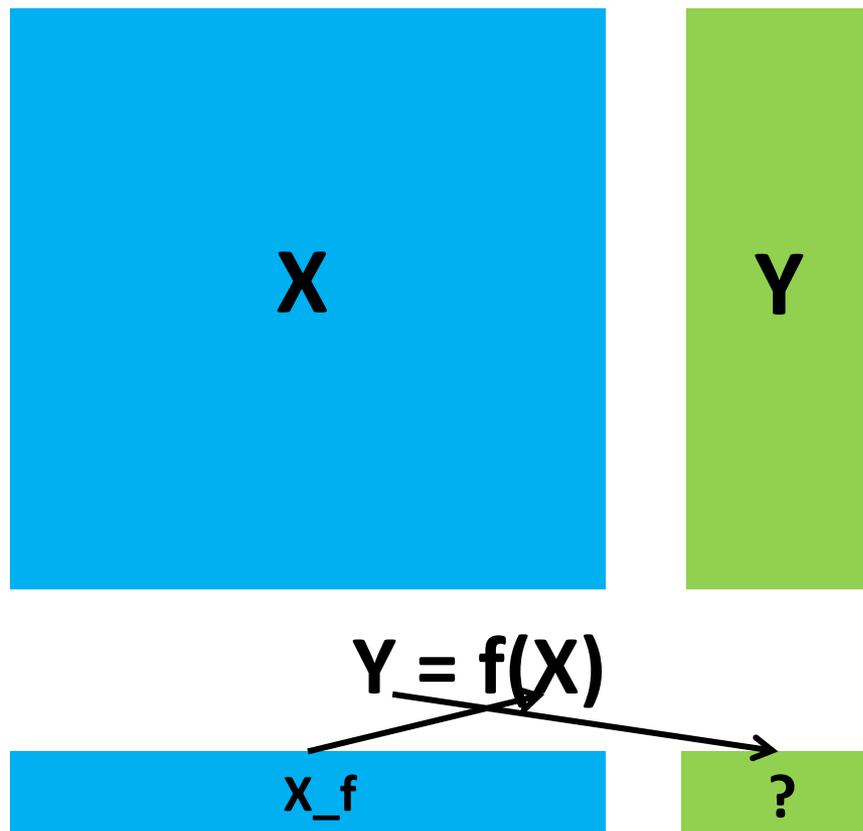
X Y

No	모델	주행거리	마력	용량	가격
1	TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	46,986	90	2,000	13,500
2	TOYOTA Corolla 1800 T SPORT VVT I 2/3-Doors	19,700	192	1,800	21,500
3	TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors	71,138	69	1,900	12,950
4	TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 2/3-Doors	31,461	192	1,800	20,950
5	TOYOTA Corolla 1.8 16V VVTLI 3DR T SPORT BNS 2/3-Doors	43,610	192	1,800	19,950
6	TOYOTA Corolla 1.6 VVTI Linea Terra Comfort 2/3-Doors	21,716	110	1,600	17,950
7	TOYOTA Corolla 1.6 16v L.SOL 2/3-Doors	25,563	110	1,600	16,750
8	TOYOTA Corolla 1.6 16V VVT I 3DR TERRA 2/3-Doors	64,359	110	1,600	16,950
9	TOYOTA Corolla 1.6 16V VVT I 3DR SOL AUT4 2/3-Doors	43,905	110	1,600	16,950
10	TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-Doors	56,349	110	1,600	15,950
11	TOYOTA Corolla 1.4 VVTI Linea Terra 2/3-Doors	9,750	97	1,400	12,950
12	TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	27,500	97	1,400	14,750
13	TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	49,059	97	1,400	13,950
14	TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	44,068	97	1,400	16,750
15	TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	46,961	97	1,400	13,950
16	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	110,404	90	2,000	16,950
17	TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	100,250	90	2,000	16,950
18	TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	84,000	90	2,000	19,000
19	TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-Doors	79,375	90	2,000	17,950
20	TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-Doors	75,048	97	1,400	15,800

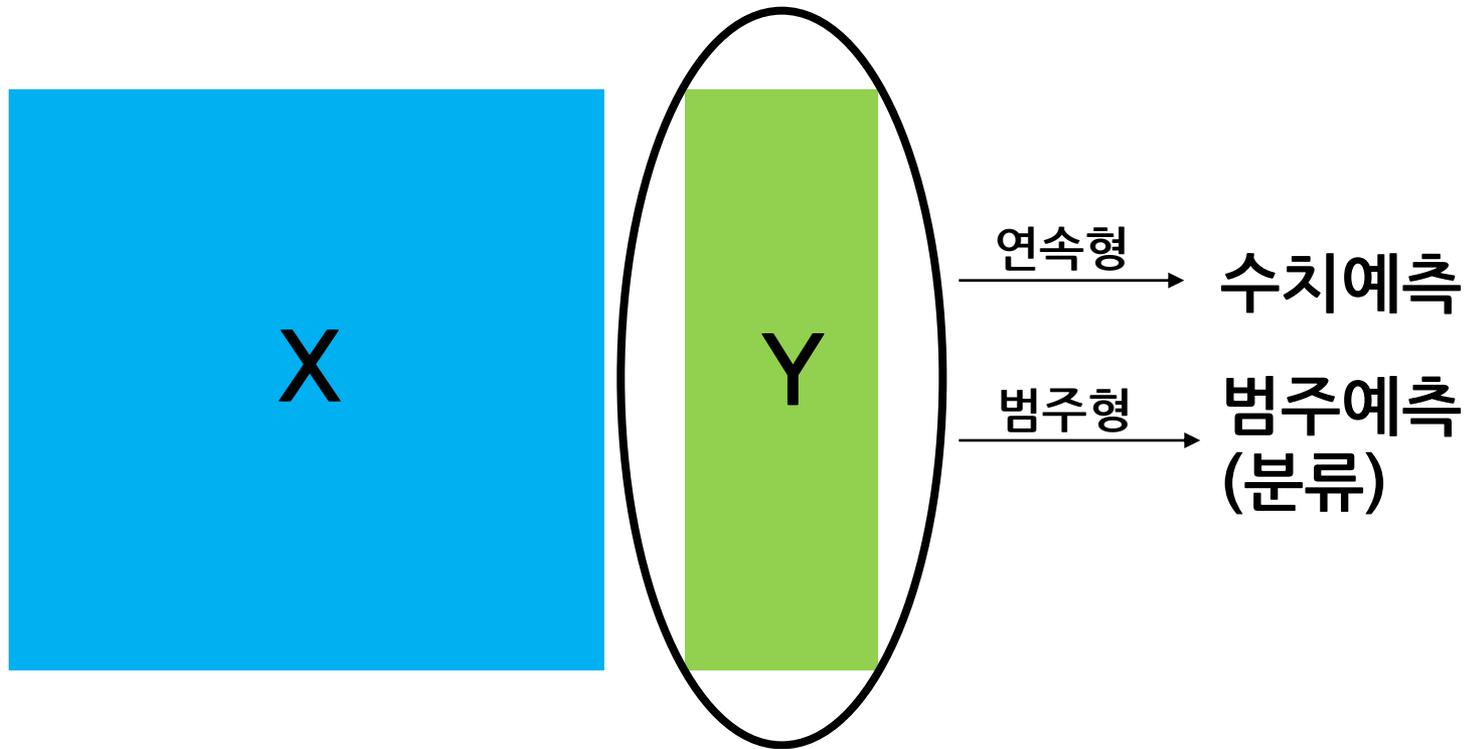
데이터마이닝 모델링?



데이터마이닝 모델링 및 예측



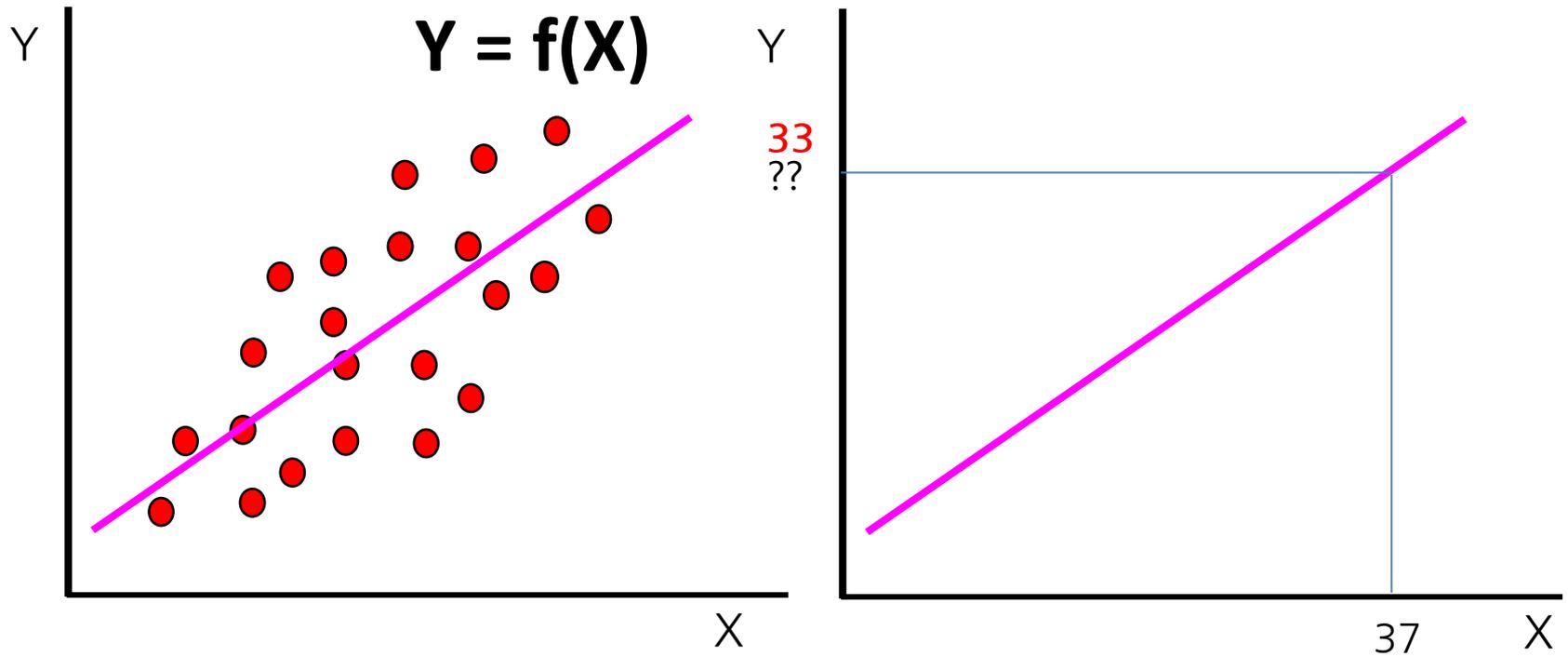
수치예측 / 범주예측 (분류)



연속형 데이터: 데이터 자체를 숫자로 표현
예) 온도, 압력, 두께, 가격, ...

범주형 데이터: 원칙적으로 숫자로 표시할 수 없는 데이터
예) 불량여부 (Y/N), 환자/정상, 이벤트 여부 (Y/N), ...

수치예측 모델링 개요

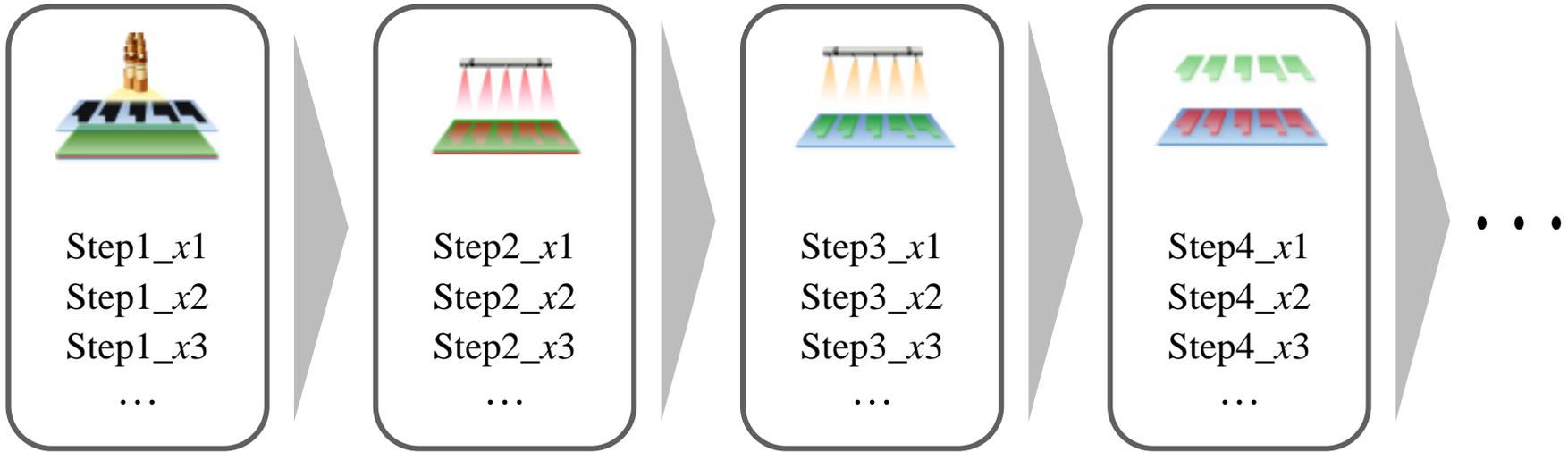


수치예측 예제 - 중고차 가격 예측

모델	X			Y
	주행거리	마력	용량 (CC)	가격
TOYOTA Corolla 2.0 D4D HATCHB TERRA 2/3-Doors	46986	90	2000	13500
TOYOTA Corolla 1800 T SPORT VVT I 2/3-Doors	19700	192	1800	21500
TOYOTA Corolla 1.9 D HATCHB TERRA 2/3-Doors	71138	69	1900	12950
TOYOTA Corolla 1.8 VVTL-i T-Sport 3-Drs 2/3-Doors	31461	192	1800	20950
TOYOTA Corolla 1.8 16V VVTI 3DR T SPORT BNS 2/3-Doors	43610	192	1800	19950
TOYOTA Corolla 1.6 VVTI Linea Terra Comfort 2/3-Doors	21716	110	1600	17950
TOYOTA Corolla 1.6 16v L.SOL 2/3-Doors	25563	110	1600	16750
TOYOTA Corolla 1.6 16V VVT I 3DR TERRA 2/3-Doors	64359	110	1600	16950
TOYOTA Corolla 1.6 16V VVT I 3DR SOL AUT4 2/3-Doors	43905	110	1600	16950
TOYOTA Corolla 1.6 16V VVT I 3DR SOL 2/3-Doors	56349	110	1600	15950
TOYOTA Corolla 1.4 VVTI Linea Terra 2/3-Doors	9750	97	1400	12950
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	27500	97	1400	14750
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	49059	97	1400	13950
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	44068	97	1400	16750
TOYOTA Corolla 1.4 16V VVT I 3DR 2/3-Doors	46961	97	1400	13950
TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	110404	90	2000	16950
TOYOTA Corolla 2.0 D4D 90 5DR TERRA COMFORT 4/5-Doors	100250	90	2000	16950
TOYOTA Corolla 2.0 D4D 90 5DR SOL 4/5-Doors	84000	90	2000	19000
TOYOTA Corolla 2.0 D4D 90 5DR TERRA 4/5-Doors	79375	90	2000	17950
TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-Doors	75048	97	1400	15800
TOYOTA Corolla 1.4 16V VVT I 5DR TERRA COMFORT 4/5-Doors	132151	110	1600	??????

수치예측 예제

Step간 연관성을 기반으로 후행 step의 공정 변수 예측



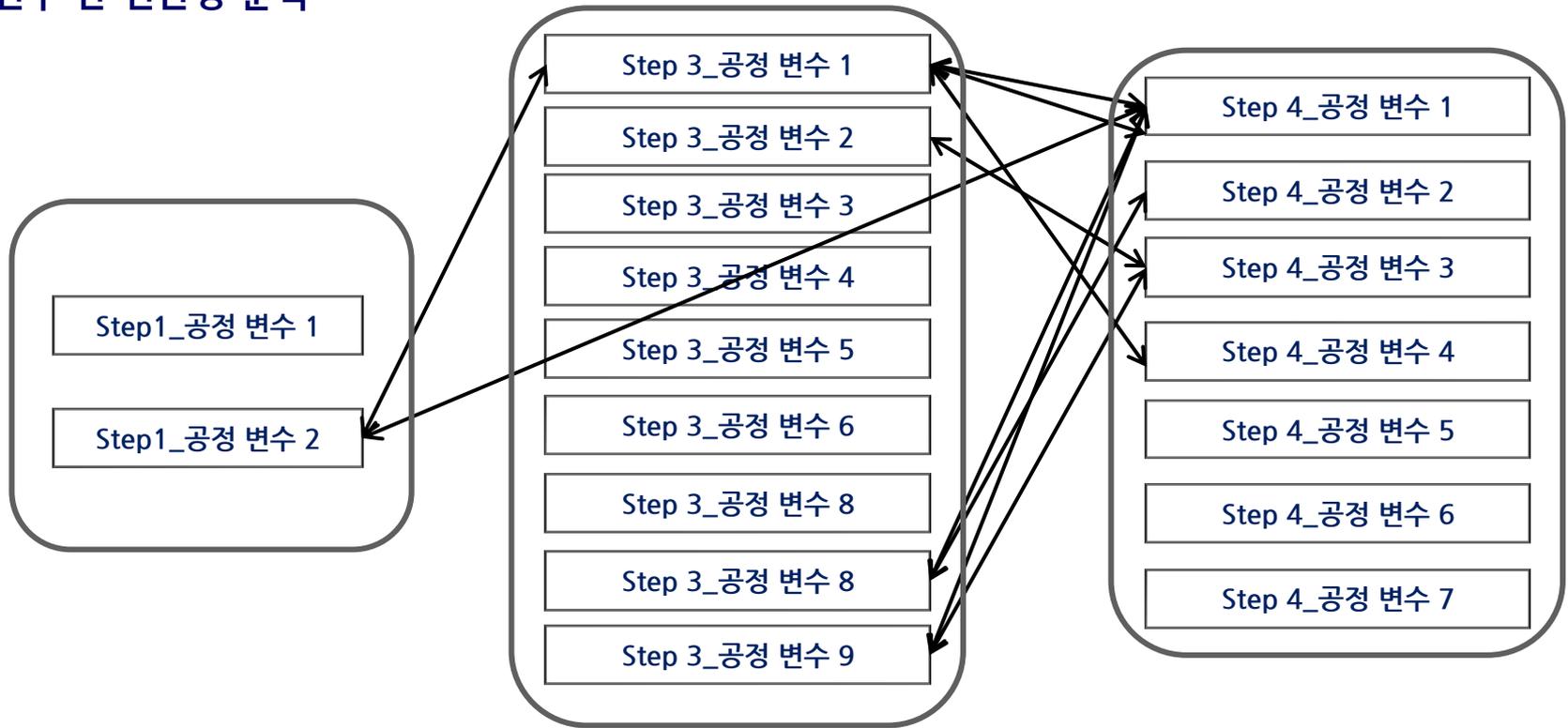
예) $Step4_x1 = f(Step1_x1, Step2_x3, Step3_x5, \dots)$

< 선행 공정변수를 활용한 후행 공정 변수 예측 모델 구축 >

- 후행 단계 공정 변수와 선행 단계 공정 변수를 연관성을 반영한 예측모델을 수립 (f)하여 후행 단계 공정 변수 값을 조기에 예측
- 데이터마이닝 기법을 활용하여 정확하고 보다 의미 있는 선/후 단계의 교호작용을 정량적으로 분석
- 이를 통해, 선행 공정단계의 변수를 조정함으로써 후행 공정 변수를 효과적으로 제어할 수 있음

수치예측 예제

공정 변수 간 연관성 분석



< 공정변수 간 연관성 맵 >

- Step 간 상관계수가 높은 공정 변수들을 연결
- 선행 공정(Step 1, 3)과 후행공정(Step4) 간에는 다소 높은 연관관계를 보임
- 이러한 연관성을 바탕으로 후행 공정의 변수의 예측모델 구축

수치예측 예제

후행 공정 변수에 대한 예측모델 구축결과 (예: LCD 공정)

선행 공정 변수(X)	후행 공정 변수(Y)	R ²
Step 1 과 3의 공정변수	Step 4_ 공정변수 1	0.79
	Step 4_ 공정변수 2	0.98
	Step 4_ 공정변수 3	0.74
	Step 4_ 공정변수 4	0.98
	Step 4_ 공정변수 5	0.12
	Step 4_ 공정변수 6	0.11
	Step 4_ 공정변수 7	0.04

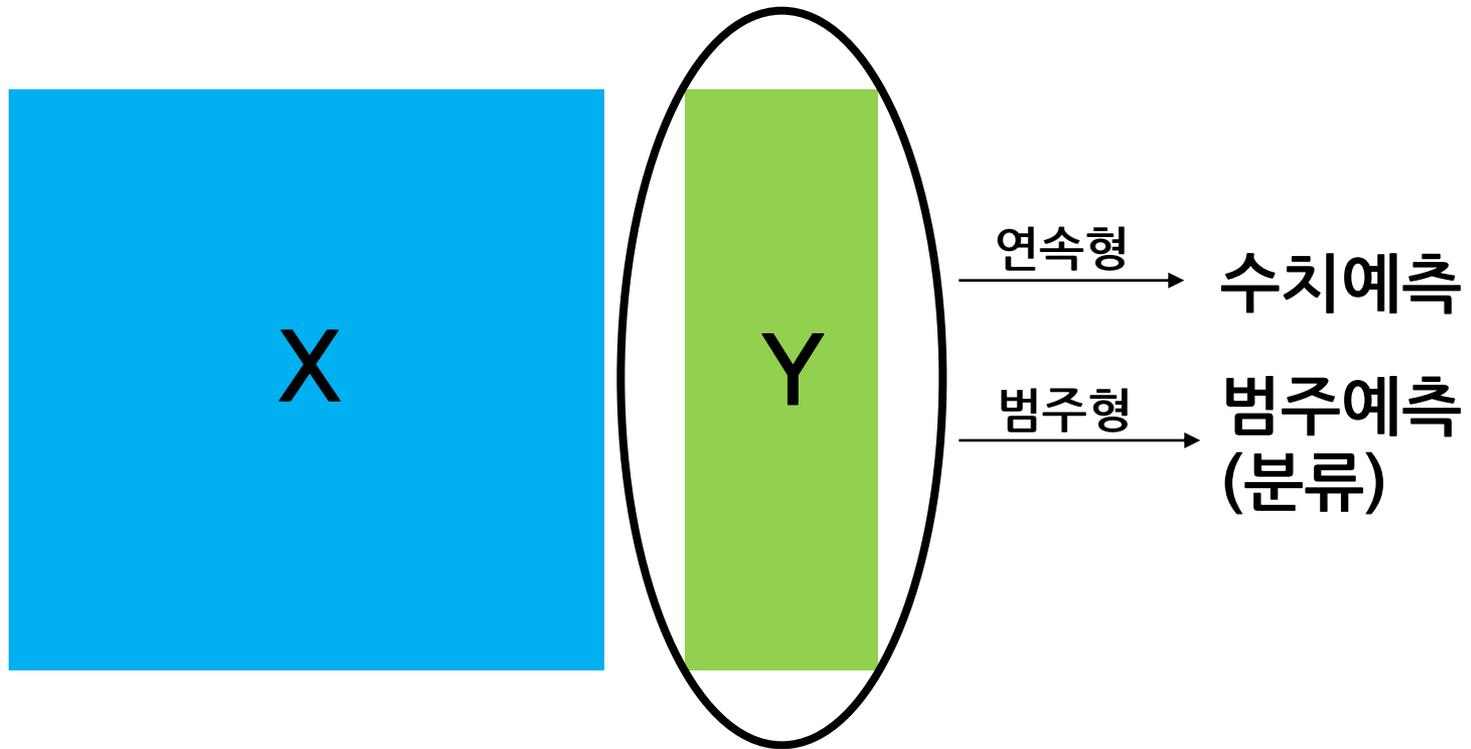
Y변수: Step4 의 공정변수

~

X변수: Step 1, 3의 전체 공정변수

- 선행공정의 변수를 독립변수(X), 후행공정의 변수를 종속변수(Y)로 설정하여 분석모델 구축
- 후행공정인 Step 4의 공정변수 중 4개의 변수(빨간색표시)는 선행 공정단계와 밀접한 관계(R² ↑)를 보임
- 설명력 (R²)이 낮은 변수들은 공정 외부 조건에 관계 있을 가능성

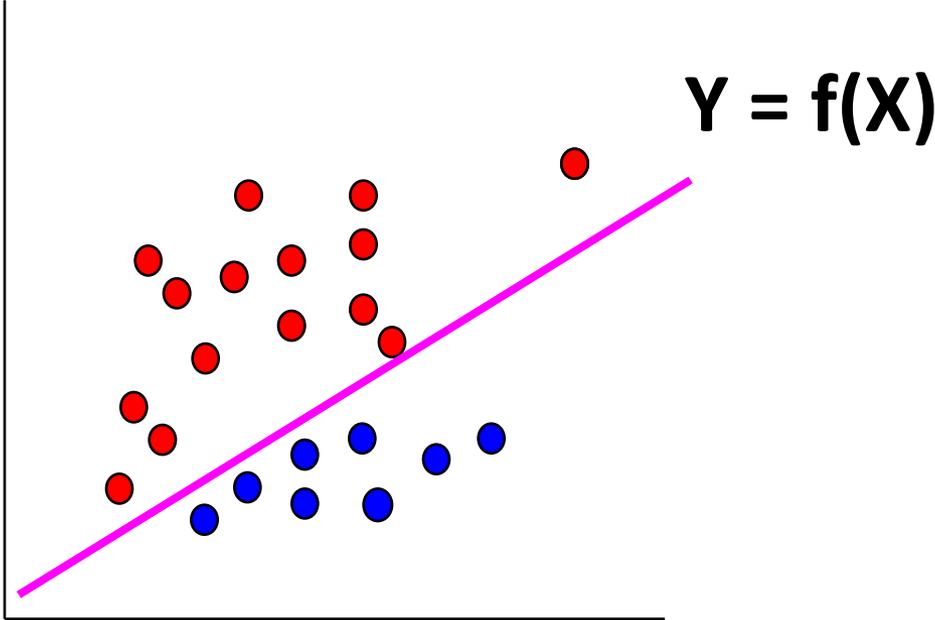
수치예측 / 범주예측 (분류)



연속형 데이터: 데이터 자체를 숫자로 표현
예) 온도, 압력, 두께, 가격, ...

범주형 데이터: 원칙적으로 숫자로 표시할 수 없는 데이터
예) 불량여부 (Y/N), 환자/정상, 이벤트 여부 (Y/N), ...

범주예측 모델링 개요



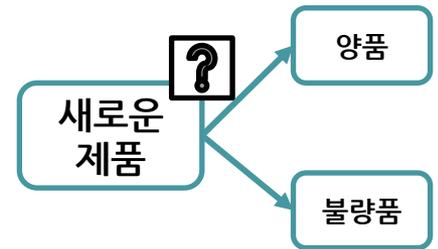
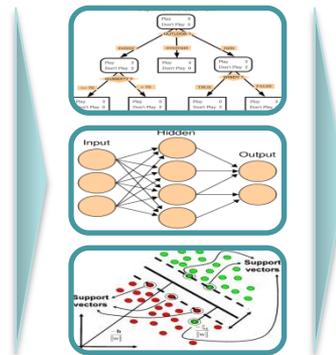
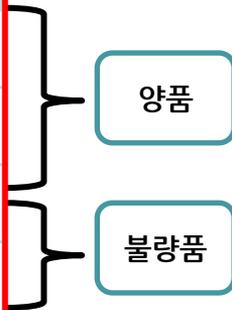
범주예측 예제 - 불량 예측

- 디스플레이 공정에서 공정 파라미터의 측정값들을 이용하여, 해당 글라스가 양품인지 불량품인지의 여부를 예측

$$f$$

공정파라미터 글라스 상태

	X_1	X_2	...	X_p	Y
제품 ₁	a_{11}	a_{12}	...	a_{1p}	1
제품 ₂	a_{21}	a_{22}	...	a_{2p}	1
⋮	⋮	⋮	⋱	⋮	⋮
제품 _n	a_{n1}	a_{n2}	...	a_{np}	0



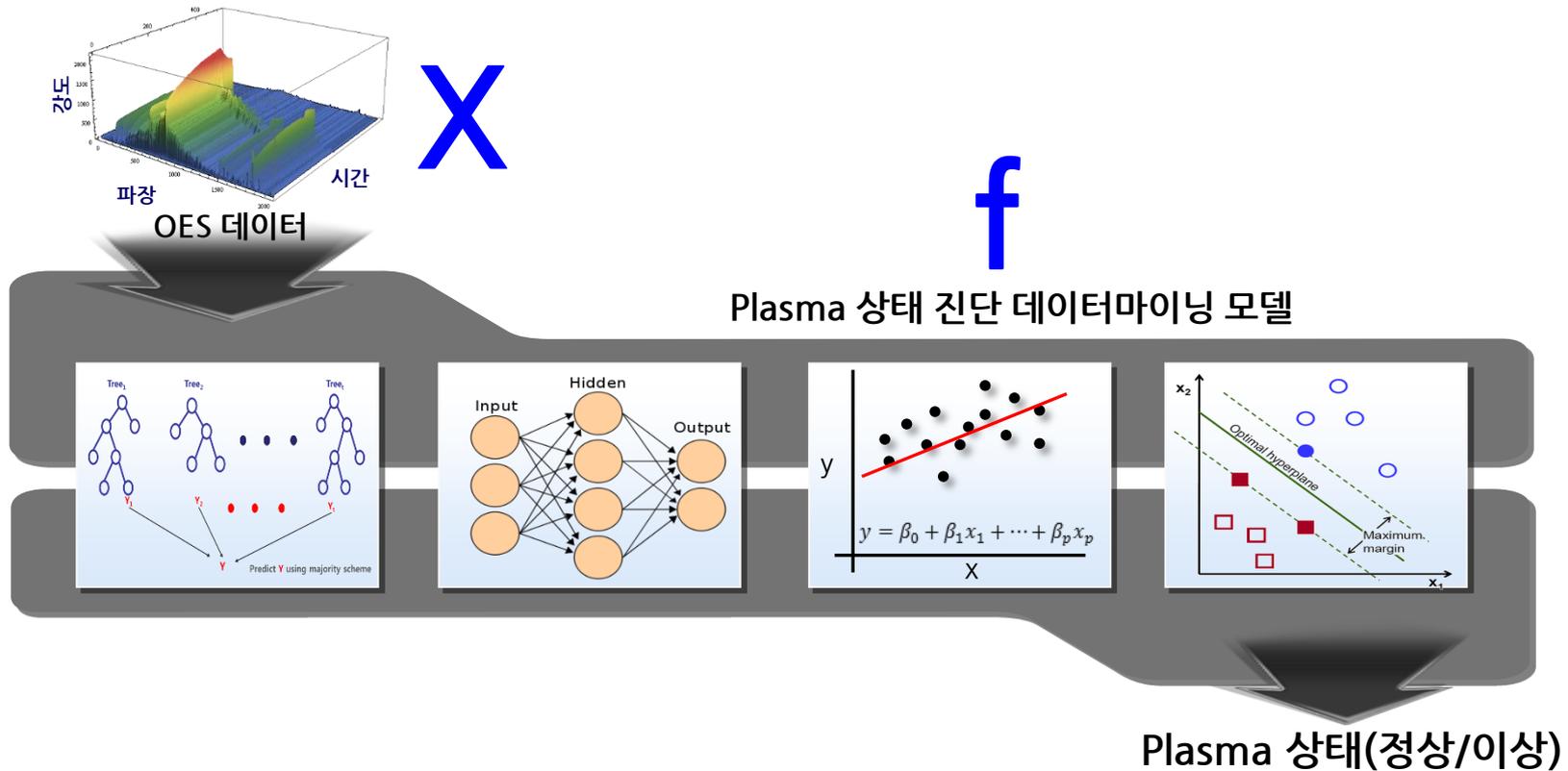
디스플레이공정 데이터

모델구축

불량 글라스 예측

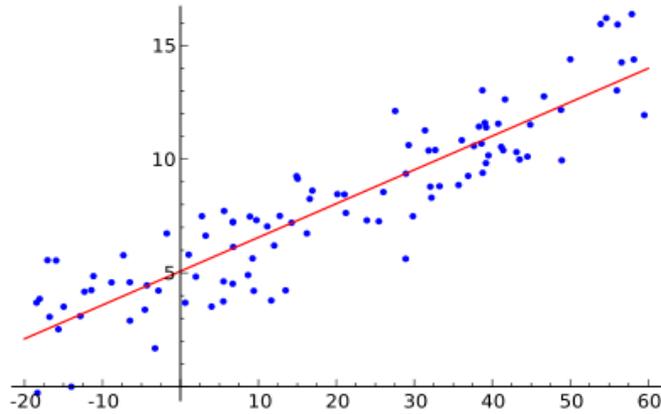
범주예측 예제 - 불량 예측

- 공정데이터를 활용하여 Plasma 이상 상태를 진단을 모델링

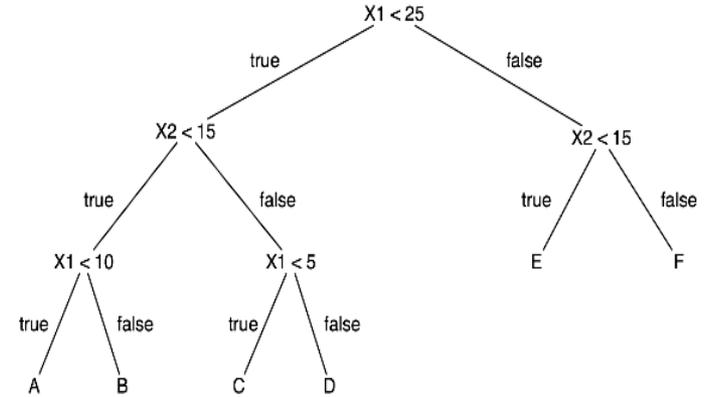


◎ 이상상태: AI으로 인한 이상, G변동으로 인한 이상

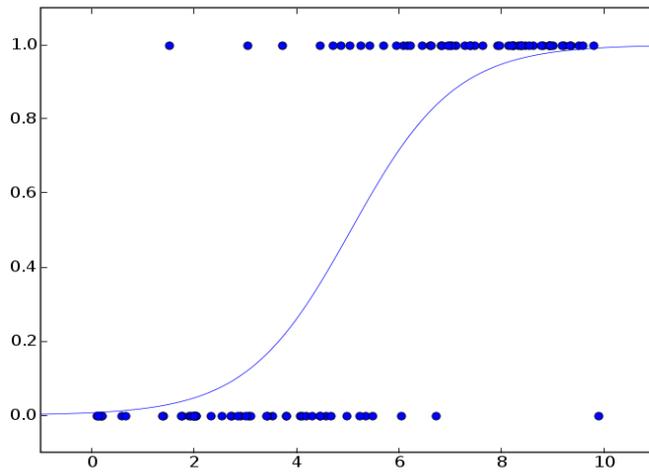
예측 알고리즘



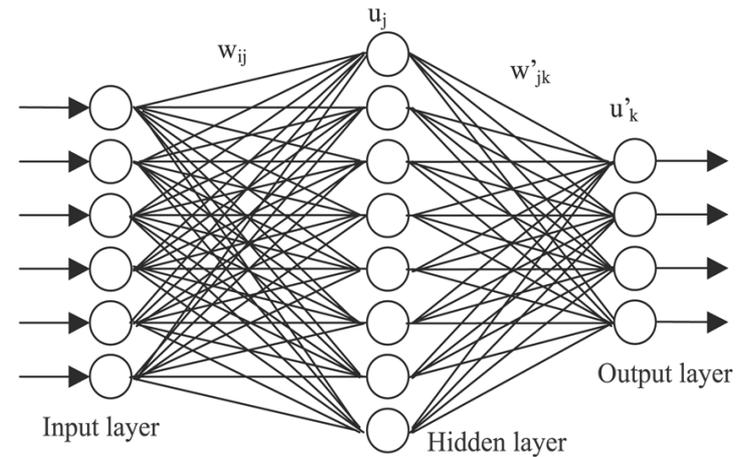
선형회귀분석



의사결정나무



로지스틱 회귀분석

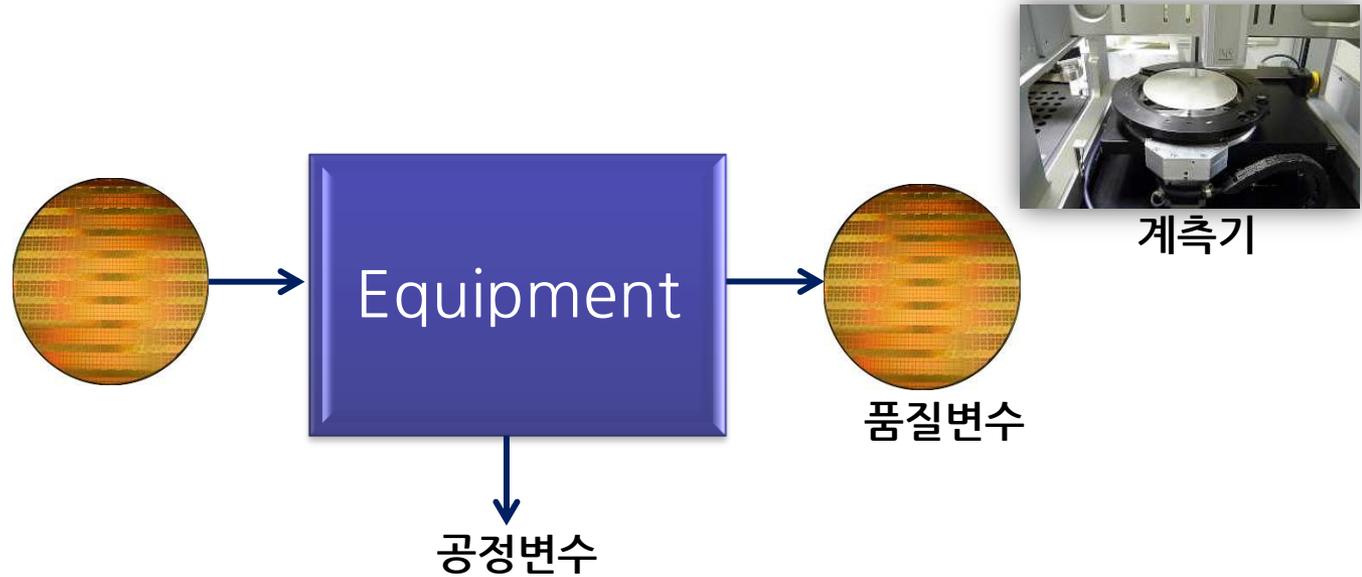


인공신경망, 딥러닝 알고리즘

- 소개
- 빅데이터 모델링 개요
- **공정 모니터링 사례**
- 빅데이터 도입에 대한 의견

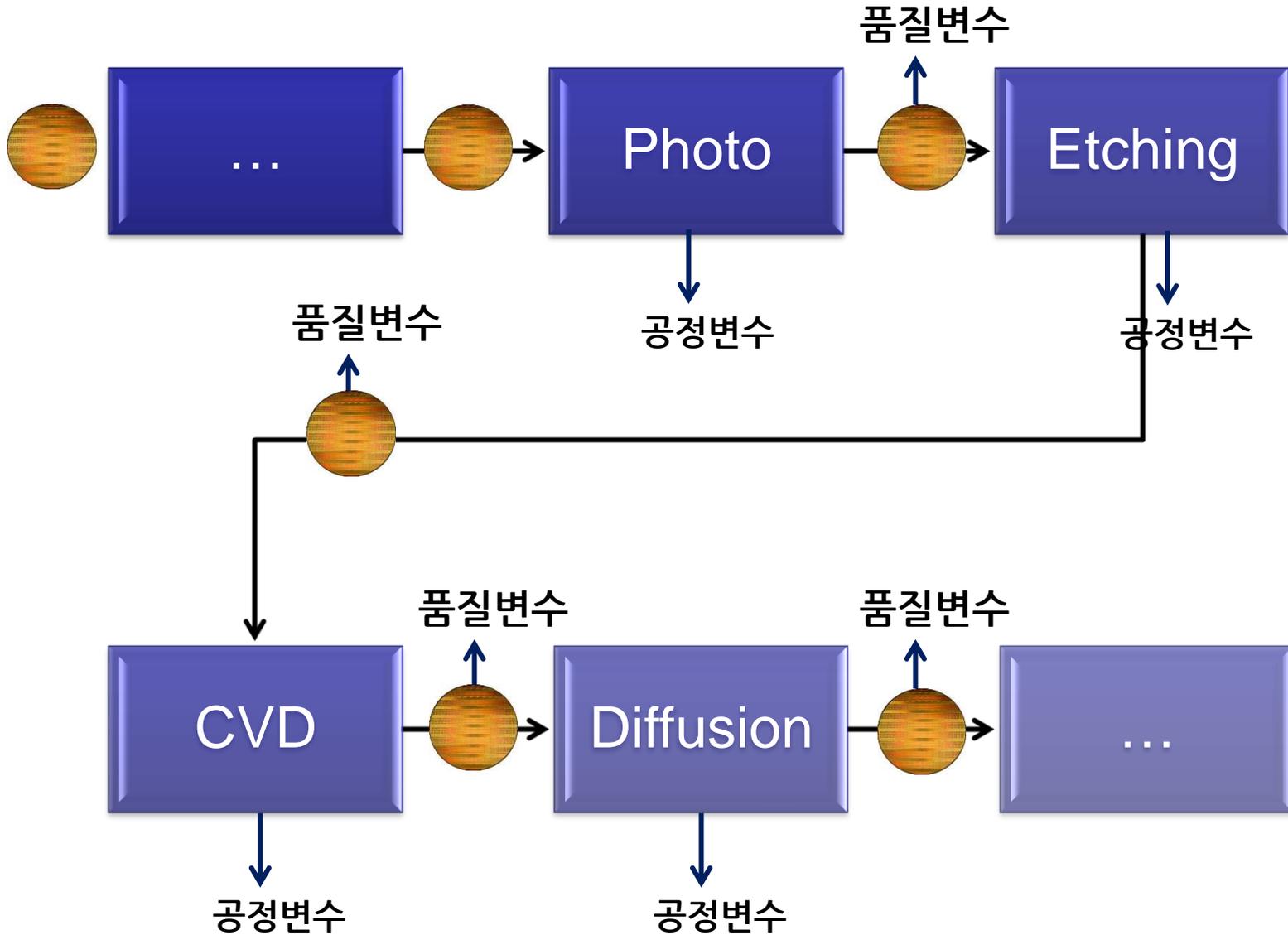
제조공정에서의 빅데이터 분석

- 수율 향상 → 궁극적 목적
- 수율 = 품질 + 생산성
- 많은 노력 → 장비개선, 재료개선, 프로세스개선, ...
- 장비와 품질계측기기 (센서)의 눈부신 발달
- 특히 고품질 달성 위해 수많은 센서 사용 → 대량의 공정변수
- 수많은 유용한 정보가 데이터 內 존재
- 수율 향상을 위한 직/간접 영향을 미치는 공정변수?
- 수율과 주요 공정변수와의 관계 규명
- 주요 공정변수의 이상변동 모니터링 및 조기탐지
- 정확한 수율을 앞선 단계에서 예측



- 공정변수: 장비의 상태를 보여주는 변수, 센서로 부터 관측 (예: power, temperature, flow rate, ...)
- 품질변수: 제품의 상태를 보여주는 변수, 계측기로 부터 관측 (예: thickness, critical dimensions, ...)

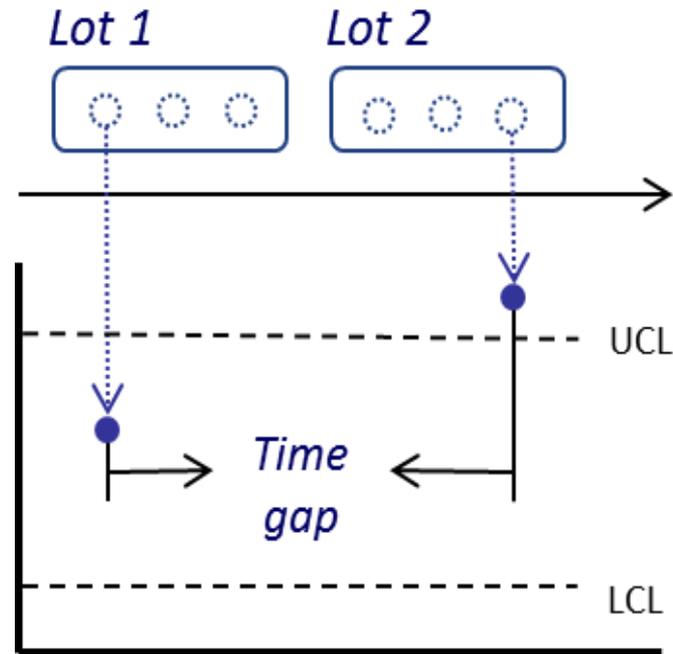
공정 프로세스



샘플링을 통한 품질검사



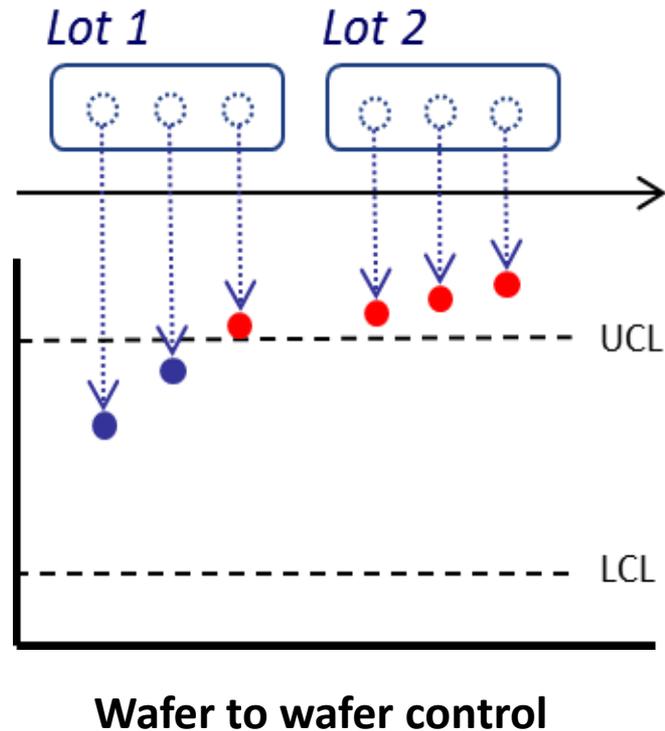
1 lot = 25 wafers



- 일반적으로 1 lot에서 1개 혹은 2개의 Wafer를 샘플링 하여 품질 검사
- Lot기반의 품질검사는 시간 차이로 인하여 불량률 조기에 탐지하기 어려움

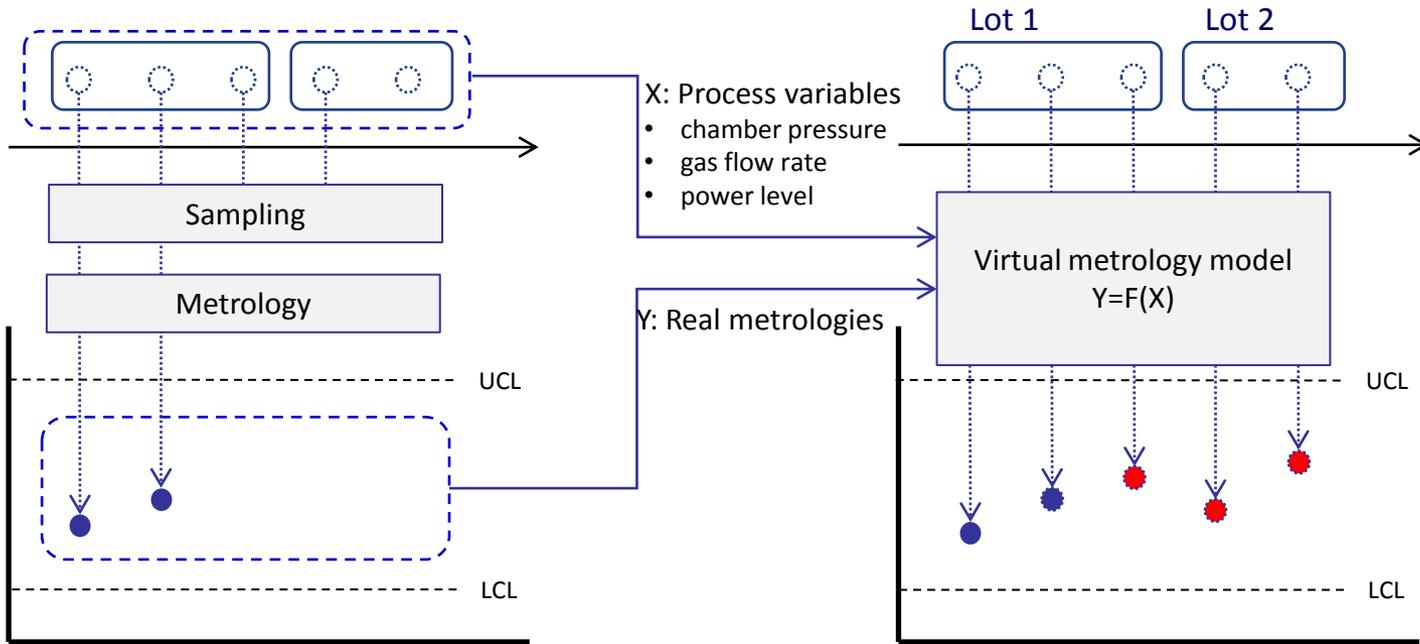
Khan, Aftab A., James R. Moyne, and Dawn M. Tilbury. "Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares." *Journal of Process Control* 18.10 (2008): 961-974.

모든 제품 검사



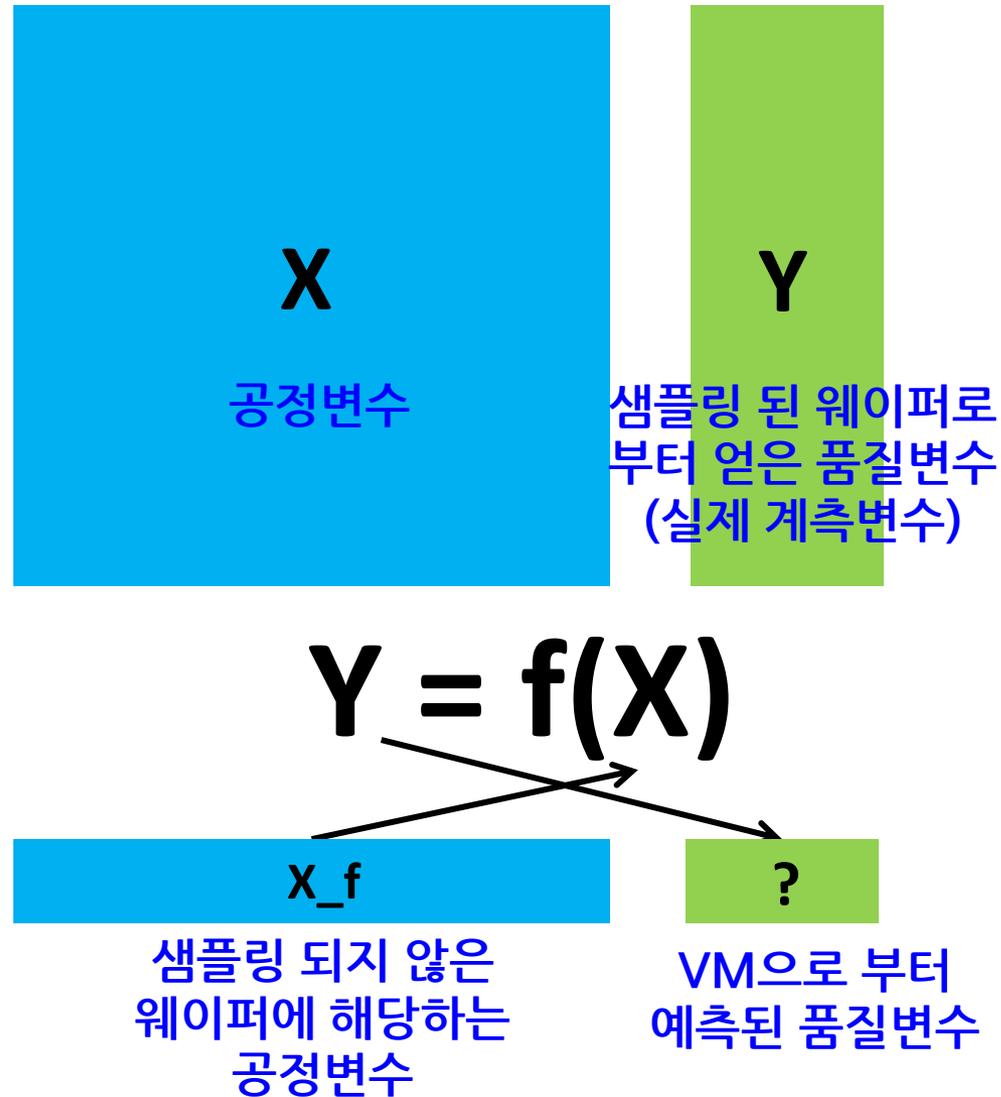
- Wafer 마다 검사 하는 것은 시간차이로 부터 기인하는 문제점을 해결
- 하지만 시간과 비용이 많이 들어 현실적으로 시행하기에 제약이 있음

가상계측



- VM 모델은 샘플링 된 웨이퍼의 품질변수와 이에 해당하는 공정 변수를 이용한 예측모델
- 구축된 VM 모델을 이용하여 샘플링 되지 않은 웨이퍼의 품질변수를 예측

Chen, PingHsu, et al. "Virtual metrology: A solution for wafer to wafer advanced process control." *Semiconductor Manufacturing*, 2005. *ISSM 2005, IEEE International Symposium on*. IEEE, 2005.



가상계측 모델링

가상계측모델

$$Y = f(X)$$

↓
품질변수

↓
공정변수

샘플링 된 웨이퍼



실제 계측기



품질변수

샘플링되지 않은 웨이퍼



+ f

가상 계측기



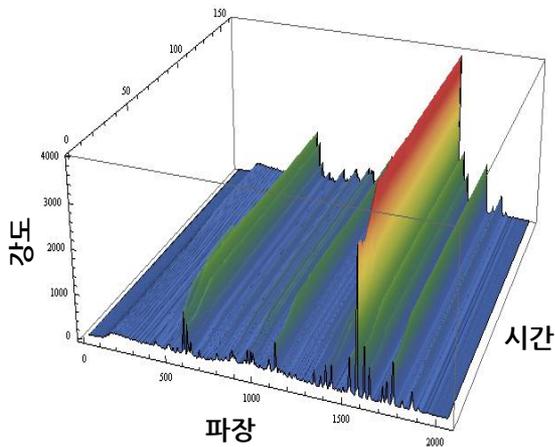
품질변수



전 웨이퍼의 품질검사

가상계측 모델링 예제

- OES (Optical Emission Signal) 데이터는 ** 공정 시 분광기 센서에서 측정된 스펙트럼 형태의 데이터임
- 독립변수 X: 분광기에서 측정된 각 시간대 파장의 강도
- 종속변수 Y: 웨이퍼의 두께
- 관측치 수에 비해 변수의 수가 매우 많은 고차원 데이터
- 인접한 변수 사이의 상관관계가 높고 중요변수들이 특정구간에 모여 있음

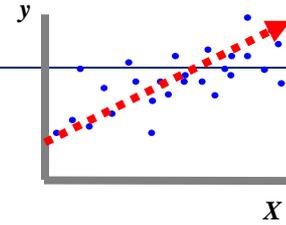
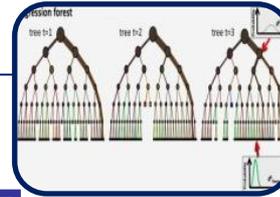


고차원 스펙트럼 데이터

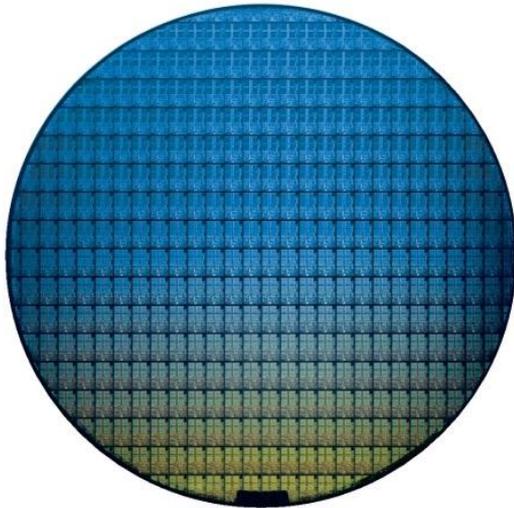
파장 \ 시간	W_1	W_2	...	W_{2049}
0.1	-2115	0	...	54
0.2	-2114	0	...	53
⋮	⋮	⋮	⋮	⋮
69.2	-2268	0	...	213



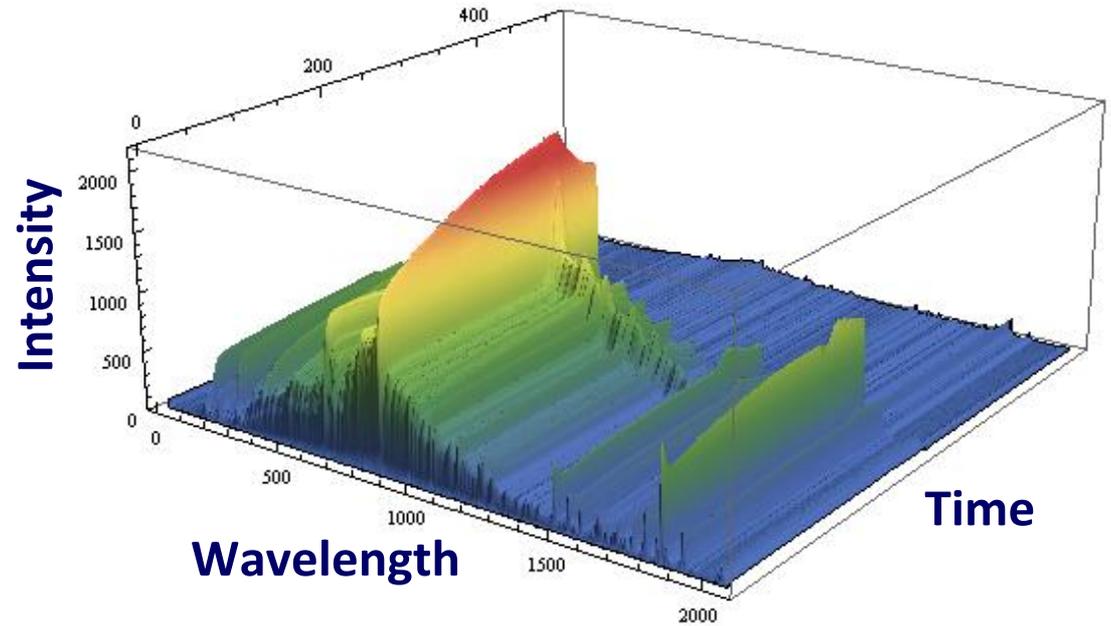
가상계측 모델링 예제



$$Y = f(X)$$

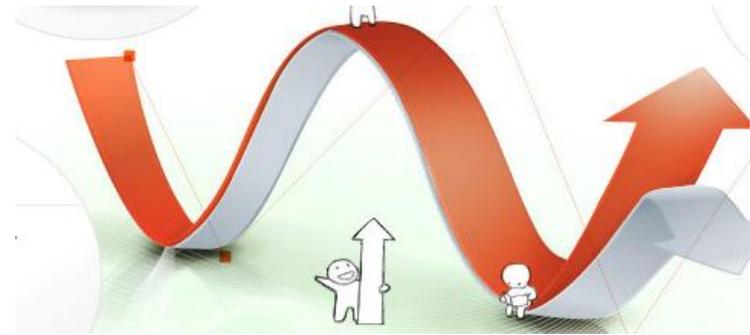


품질변수

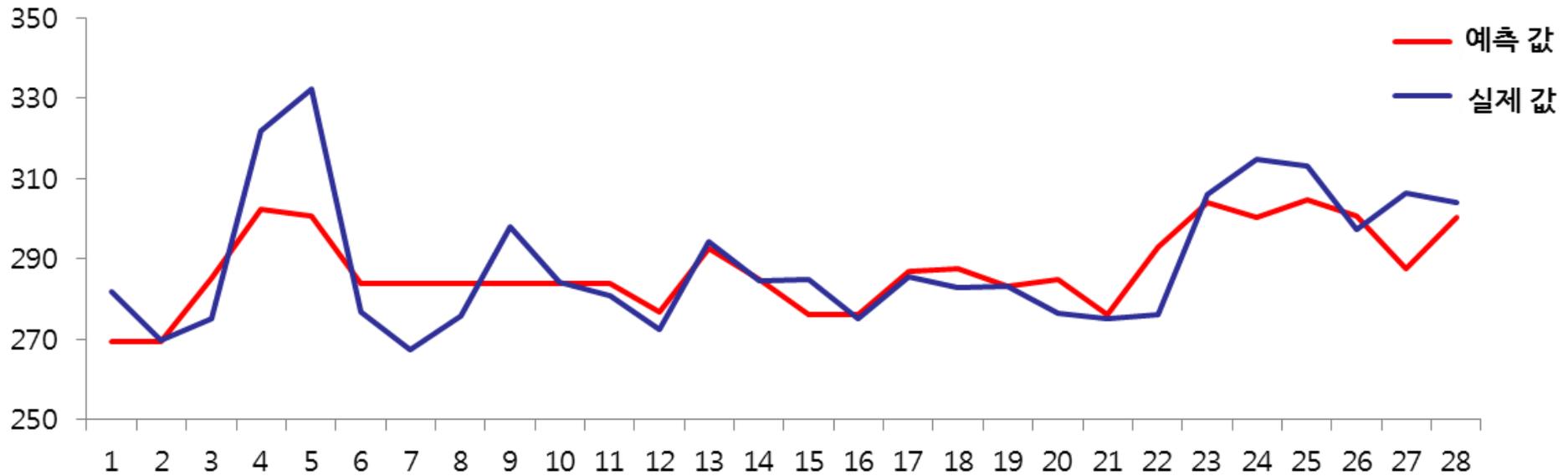


공정변수: OES 시그널

가상계측 모델링 예측 결과



<Fused Lasso 기반의 변수선택을 적용한 k -NN 회귀모델>



다변량 공정 모니터링

단변량 기반 설비/공정 이상에 대한 감지 활동을 진행 하고 있으나 관리적인 한계와 특정 이상에 대한 검출력 저하로 새로운 공정 관리 방법이 필요

단변량 관리도의 급격한 증가

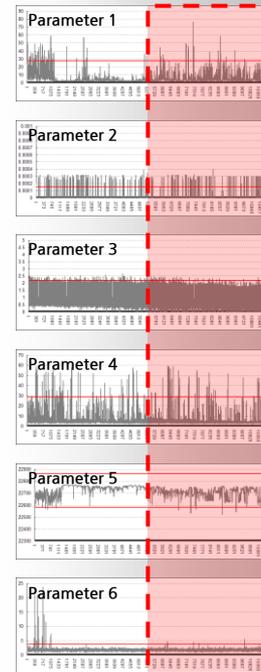
▪ FDC 관리도 증가 추이 (1개 FAB 기준)



- ✓ 관리도 Spec관리 시간의 증가
- ✓ FDC False Alarm 건수의 증가

단변량 관리도의 검출력 한계

▪ 공정/설비 이상 발생 시 FDC Alarm 항목

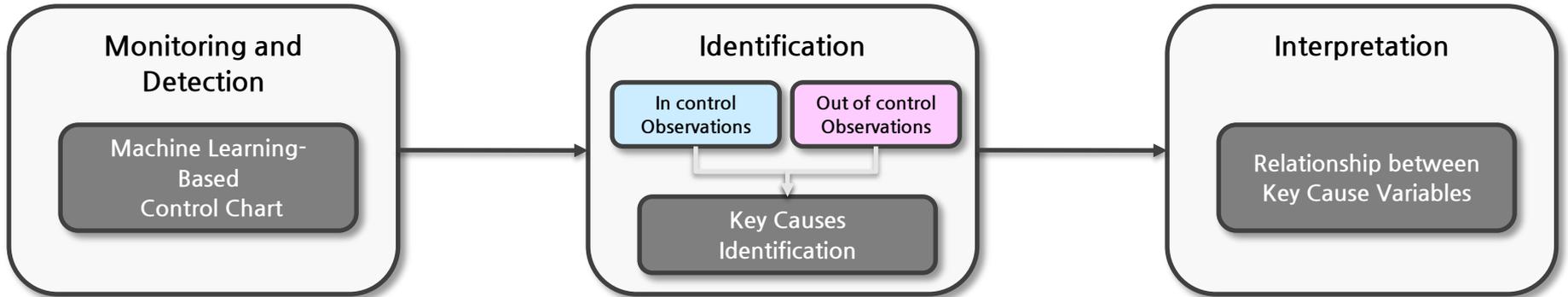


이상구간

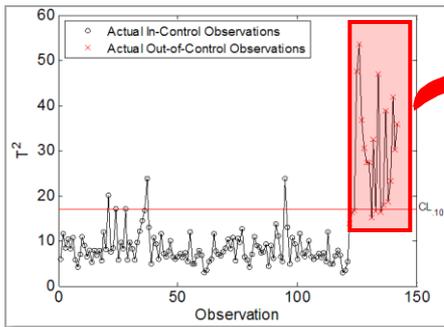


- ✓ 공정/설비 이상 발생 시 FDC Alarm에 대한 True/False 분석이 어려움
- ✓ 공정/설비 이상의 71%가 복수 항목에서 FDC Alarm 발생 (평균 6.8개 항목)
- ✓ 공정별 Engineer 1인 기준 4.6hr/일 FDC Alarm 조치

단변량 관리 대비 우수성이 검증된 다변량 관리에 대하여 관리도의 이탈부터 최종 원인항목 분석까지 일련의 분석 프로세스를 3단계로 정의

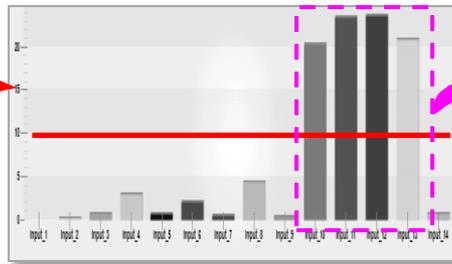


▪ 머신러닝 기반 관리도



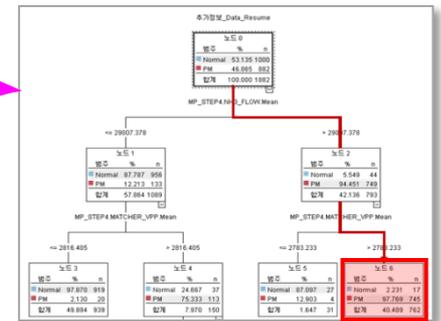
: 실시간 공정 이상 모니터링 수행

▪ T² Decomposition 결과 분석



: 이상 원인 분석 기법을 활용한 이상을 발생 시키는데 기여한 원인 Parameter 선정

▪ Decision Tree 분석



: 이상 원인 Parameter간 교호작용 및 선후관계를 해석

- 소개
- 빅데이터 모델링 개요
- 공정 모니터링 사례
- 빅데이터 도입에 대한 의견

누구도 빅데이터 도입의 효과를 확신하지 못하고 있는 지금 ...



위에선 빅데이터를 도입하라.

말처럼 쉬운 일이 아니다.

경험과 직관에 의존했던 문화를 바꾸는 것은 쉽지 않아

현재 직관과 경험에서 나온 의사결정을 확인하는 용도로 부터 시작해야

과거실적의 비판으로 사용되어서 안되고

기존에 없었던 완전히 새로운 정보를 찾으려는 용도로 사용돼서도 안돼

모든 공정 관심 문제를 빅데이터 기법으로 해결할 수 있나?

NO!

1. 공정 불량 원인을 제공하는 데이터 자체가 수집 안 되고 있어
2. 적절한 기법을 사용하지 않고 있어

1. 공정 불량 원인을 제공하는 데이터 자체가 수집 안 되고 있어

- 되는 것 부터 하자
- 하면서 무엇이 필요한지 파악하자

2. 적절한 데이터마이닝 기법을 사용하지 않고 있어

- 해당 문제의 도메인 지식이 있는 분석가 필요
- 마케터/엔지니어의 인풋이 필수

빅데이터로 성공하려면

1. 의사결정 목표를 확실히 세우고 문제점을 정확히 파악
2. 외부데이터 (트위터, 블로그, ...) 보다는 내부데이터 분석에 초점
3. 적절한 분석기법
4. 분석결과로 부터 의미 있는 인사이트를 도출
5. 성공 케이스 확산
6. 경험과 감에 의존하는 분위기에서 좀더 객관적인 근거에 의존하는 분위기로 바뀌어

감사합니다