



GS칼텍스

미래비리더  
MAKING SOMETHING GOOD



# G-Yes

AI 기반의  
폴리머 물성 예측 알고리즘을 통한  
소재 개발서비스 플랫폼

팀원: 김나현 신충현 황미경 황정훈  
지도교수: 전종식 교수

## G-Yes



신충현



황미경



황정훈



김나현

# 목 차

## 1. 주제 선정 및 배경

- 01.폴리머란?
- 02.폴리머 시장
- 03.연구개발의 한계
- 04.주제 선정

## 2. 데이터 분석

- 01.데이터 정의
- 02.데이터 문제
- 03.데이터 분석 scheme
- 04.데이터 전처리

## 3. 알고리즘

- 01.알고리즘 비교
- 02.알고리즘 선택
- 03.Algorithm Summary

## 4. 추가 서비스

- 01.Fault Detection Alarming
- 02.웹 서비스 배경영상

## 5. 기대효과

- 01.웹서비스
- 02.Customizing



# 1. 주제 선정 및 배경



# 01. 폴리에탄?

1 주제 선정 및 배경



폴리프로필렌, 폴리아미드 등의 주 소재에 충전제와 첨가제를 더해 사출, 압출하거나 열을 가해 가공하여 만드는 제품

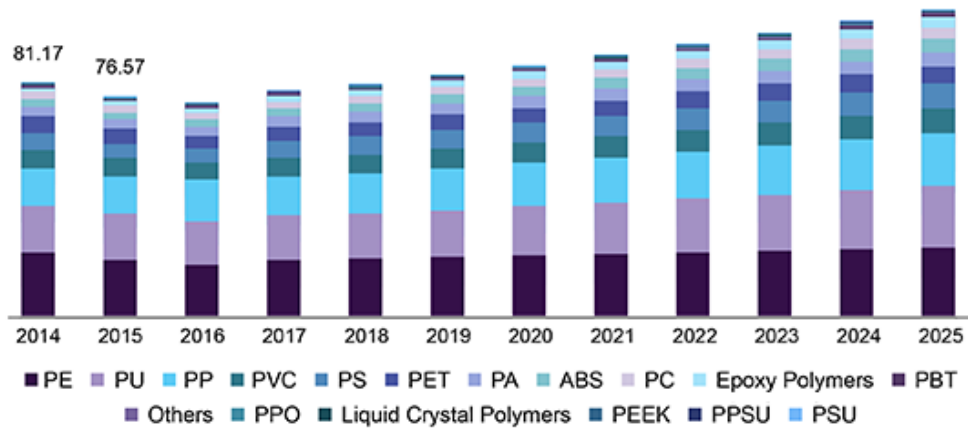
## 폴리프로필렌 사용제품



# 02. 폴리에터 시장

1 주제 선정 및 배경

U.S. plastics market size, by product, 2014 - 2025 (USD Billion)



Source: www.grandviewresearch.com

세계 플라스틱 시장 규모가 2017년 5,226억 달러에서 2025년까지 연 평균 4.0%로 성장 할 것으로 전망 .  
건설, 자동차, 전기 전자 산업에서의 플라스틱의 소비가 예측 기간 동안의 성장을 뒷받침함.

플라스틱 시장의 성장

대량 생산 및 공급 요구

고기능성 소재 요구

복합수지 신소재 개발 요구

# 03. 연구개발의 한계

1 주제 선정 및 배경



## 비용

개발에 대한 위험부담이 클 뿐만 아니라 많은 연구 개발 비용 발생.



## 시간

신소재를 연구 및 개발하는 데에 오랜 시간이 소요.



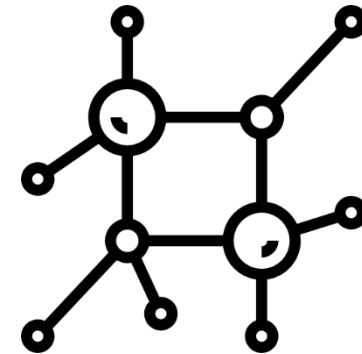
## 인원부족

연구와 영업에 필요한 인원이 부족.



## 알고리즘

3가지 한계를 보완할 최적화 레시피 알고리즘 필요.



# 04. 주제 선정

1 주제 선정 및 배경



WHAT?

Input 물질에 대한 정보로  
생산물 특성 예측 알고리즘 개발



HOW?

개발 레시피에 대한  
결과 특성치를  
미리 예측



WHY?

연구 개발 시  
비용과 시간 절감 필요,  
신소재 시장 내 수요 증가





## 2. 데이터 분석

# 01. 데이터 정의 : X값

2 데이터 분석

Feature 90개

데이터 811개	PP		...	Filler		...	Rubber		...	Others	
	P001	...	...	F001	...	...	R001	...	...	OTH1	...
			...			...			...		
			...			...			...		
	.	.		.	.		.	.		.	.
	.	.		.	.		.	.		.	.
	.	.		.	.		.	.		.	.
	.	.		.	.		.	.		.	.
	.	.		.	.		.	.		.	.
			...			...			...		

Polypropylene: 41개

Filler : 18개

Rubber : 22개

Others : 9개

총 90개의 feature

811 번의 실험 데이터

# 01. 데이터 정의 : Y값

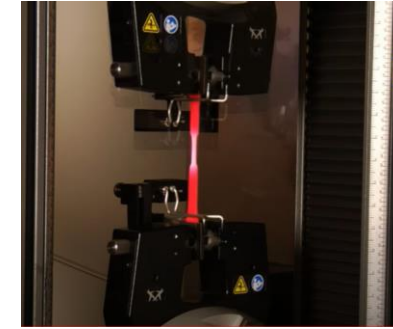
2 데이터 분석



비중



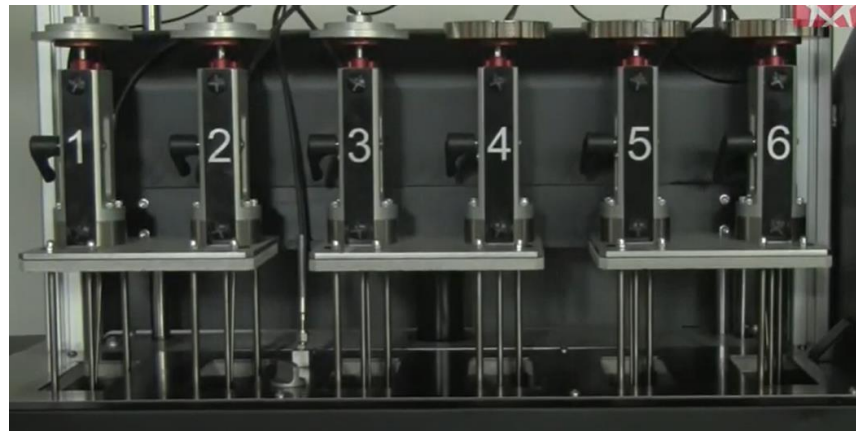
MI



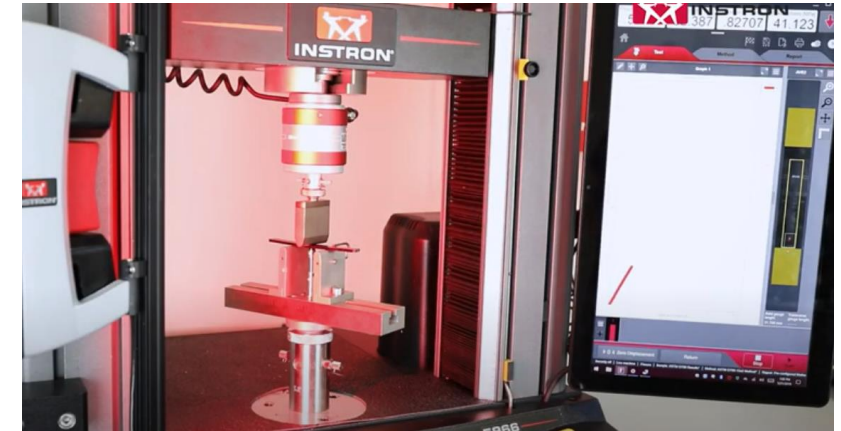
인장강도



IZOD 충격강도



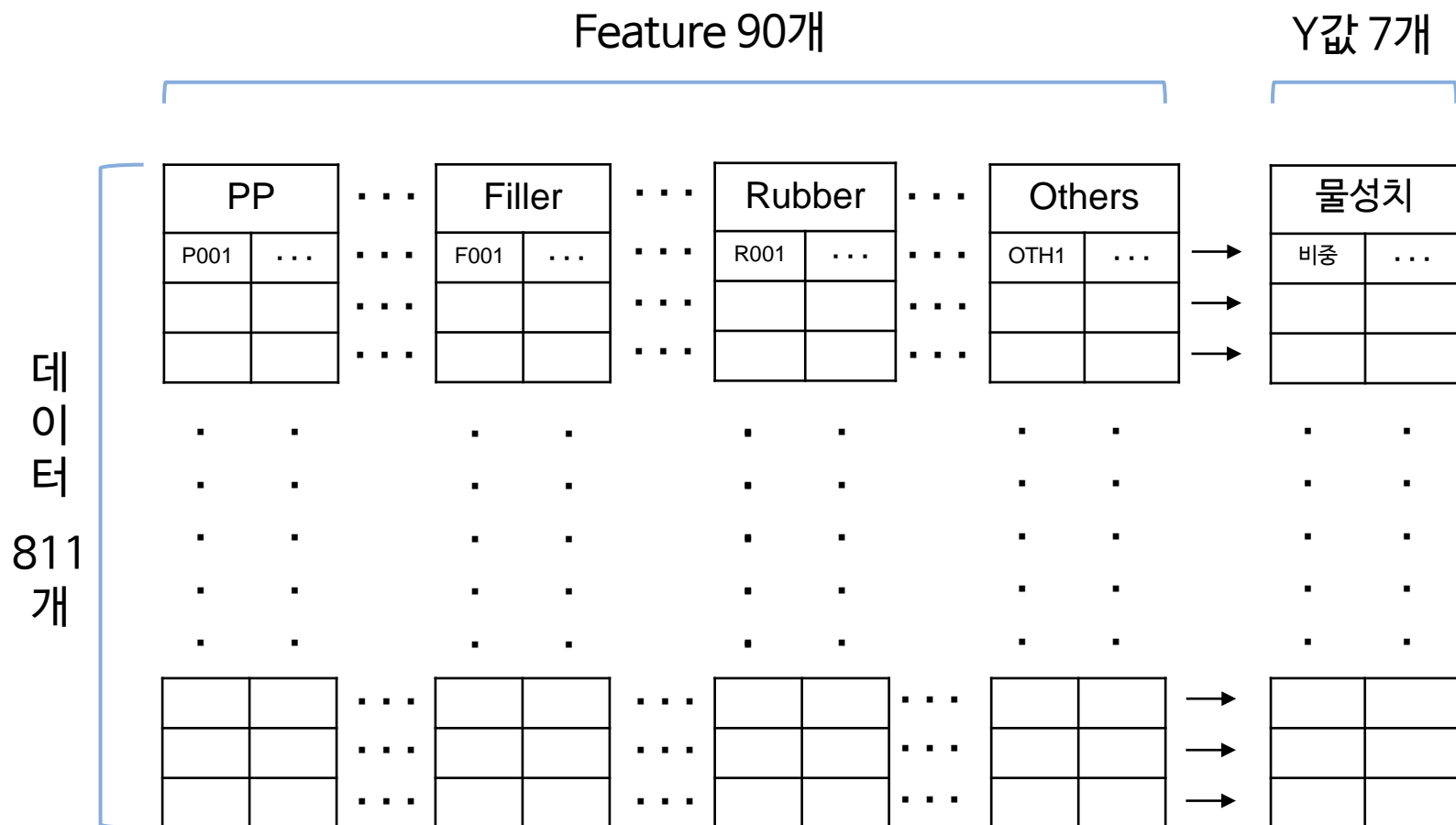
HDT



굴곡 탄성률, 굴곡 강도

# 02. 데이터 문제

2 데이터 분석



High-Dimension (>90D) & null

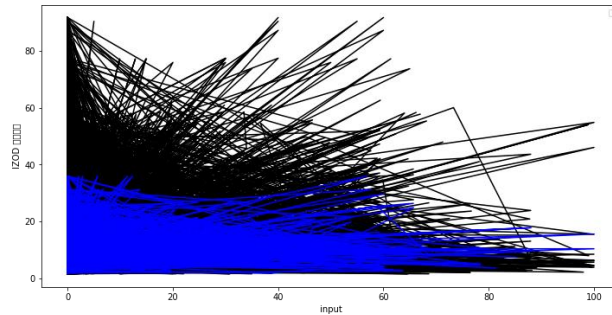
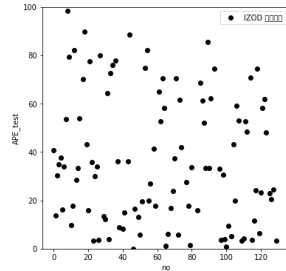
같은 input에 다른 output  
→ 통일성 X

결측치, 이상치

Data 수가 너무 적음,  
Y값이 7개

# 02. 데이터 문제

2 데이터 분석



처음 주어진 유량 데이터만으로는  
accuracy 향상 한계

문제 해결

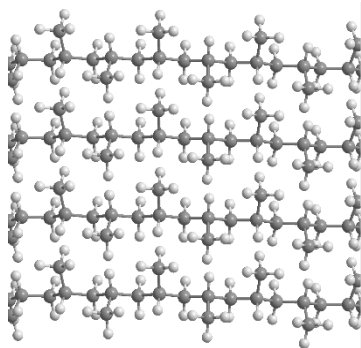


데이터 전처리,  
Algorithm 선택

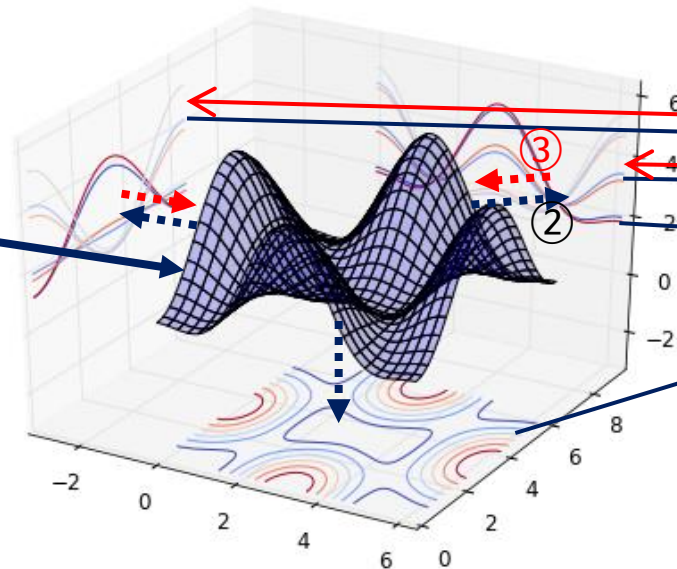
# 03. 데이터 분석 Scheme

2 데이터 분석

Interaction between Atoms (C,H)  $\xrightarrow{\text{mapping}}$  molecular configuration  $\xrightarrow{\text{mapping}}$  property



①



Property =  $f(\text{potential } E)$

- MI,  $P_1 = M_1 * F$
  - Force Module,  $P_1 = M_2 * F$
  - 굴곡 탄성률,  $P_3 = M_3 * F$
  - 충격 강도,  $P_4 = M_4 * F$
- ( $F$ =force field matrix,  
 $M_i$ =mapping matrix ( $i=1 \sim 4$ ))



# 04. 데이터 전처리

2 데이터 분석

구분	Density	FM	HDT	Izod	MI	TS	data-구
P001	0.90	1450	90	4	3.5	34	43
P002	0.90	1550	90	3.5	11	35	59
P003	0.90	1550	90	3.5	25	35	70
P004	0.90	1550	95	6.5	0.7	35	48
P005	0.90	1250	85	8	3	26	40
P006	0.90	1250	85	7	9	26	98
P007	0.90	1250	85	7	9	26	40
P008	0.90	1500	105	5	27	28	83
P009	0.90	1500	105	5	45	28	48
P010	0.90	1450	105	4.5	60	28	41
P011	0.90	1600	105	50	0.6	29	52
P012	0.90	1506.414	110.2559	5.49286	30	29.37429	43
P013	0.90			1500			42
P014	0.90	706.1315	70.46433	N/B	21	16.52304	80
P015	0.90	947.3578	78.75425	N/B	11	17.44098	40
P016		1600.565	111.9139	9.154766	10	30.29223	59
P017		1788.867	111.9139	5.950598	60	30.29223	70
P018					61		62
P019		1788.867	111.9139	6.408336	30	30.29223	197
P020		941.5087	91.18913	45.77383	28	19.27688	42
P021					30		48
P022		1788.867	111.9139	5.49286	100	30.29223	131
P023	0.91	1977.168	116.0589	3.661906	6	38.55375	41
P024	0.90	611.9807		N/B	20	12.85125	63
P025	0.91	1977.168	116.0589	3.661906	14	38.55375	48
P026	0.91	1977.168	116.0589	2.74643	45	38.55375	93
P027	0.90	2412.263	91.18913	1.830953	40	32.12813	58
P028	0.90	706.1315	74.60929	N/B	40	16.52304	88
P029	0.91	1600.565	111.9139	9.154766	10	30.29223	173
P030	0.91	1694.716	111.9139	6.408336	30	30.29223	269
P031	0.90	1035.66	82.89921	13.73215	9	24.78455	42
P032	0.90	1289.867	99.47905	7.323813	30	25.7025	53
P033	0.90	1506.414	91.18913	3.661906	14	31.21018	42
P034	0.91	1694.716	107.769	5.49286	100	30.29223	207
P035	0.90	1223.961	99.47905	7.323813	52	24.78455	52
P036	0.90	1271.037	103.624	3.661906	110	22.03071	48
P037		1223.961		7.323813	35.5		41
P038		1223.961		8.239289	39		43
P039							66
P040							40
P041							40

1. 결측치 최소화 고려
2. Independence 고려

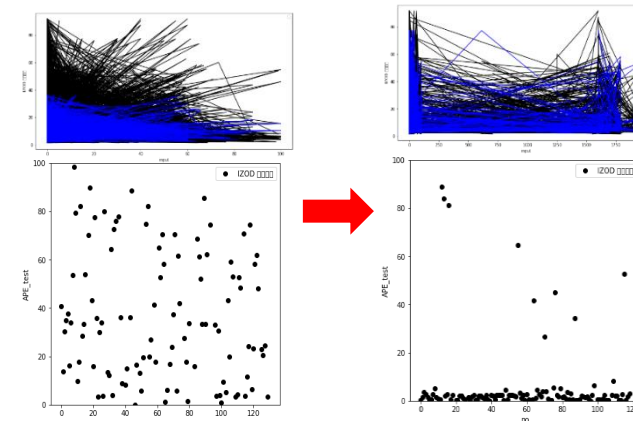
## 결과

91D



<42D

1. 차원 축소
2. P001~P041 외 새로운 input 예측 가능



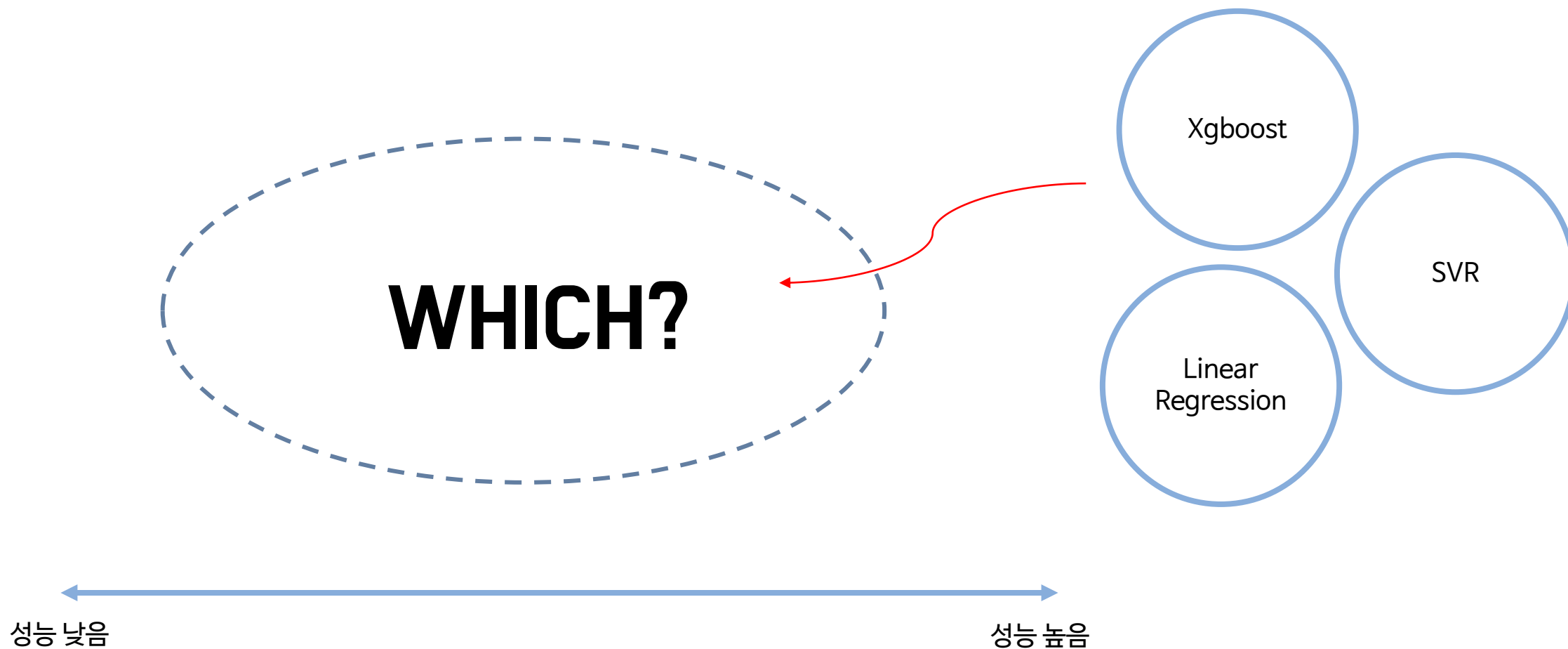
Model accuracy 향상



## 3. 알고리즘

# 01. Algorithm 비교

3 알고리즘



# 02. Algorithm Summary

3 알고리즘

## Linear Regression

- 선형관계인 경우 효율적으로 계산이 쉬워지고, 과적합이 적음

## XG boost

- boosting 방식의 ensemble 기법으로 여러 개의 약한 학습기를 순차적으로 학습-예측하면서 잘못 예측한 데이터에 가중치 부여를 통해 오류를 개선
- 이때, 가중치 업데이트 과정에서 경사하강법(Gradient Descent)를 이용
- gradient boosting model의 overfitting 문제를 hyperparameter 조정을 통해 최적화

## SVR

- 적은 data (<800) → 화학적 결합력의 확률적 분포 확인 어려움  
⇒ Margin 최대화를 통해 분포 공간 split
- 비선형 system + 적은 data 수 → 낮은 정합성  
⇒ Kernel function을 이용한 비선형 모델링 + Least squares 계산을 통한 이론적 기반 계산

Linear

Nonlinear

비중

굴곡 탄성률

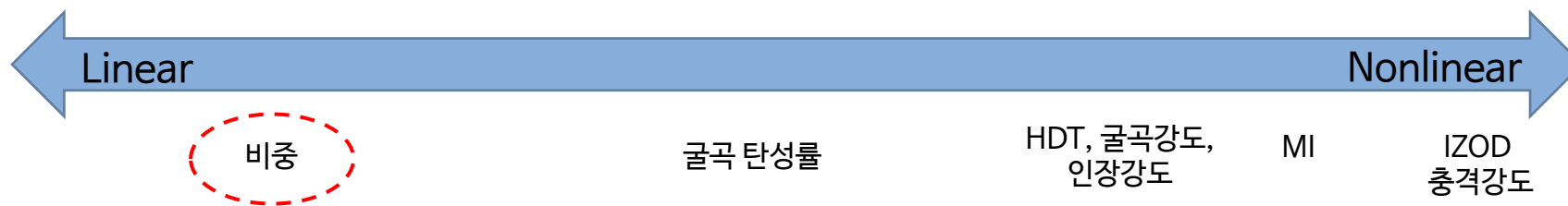
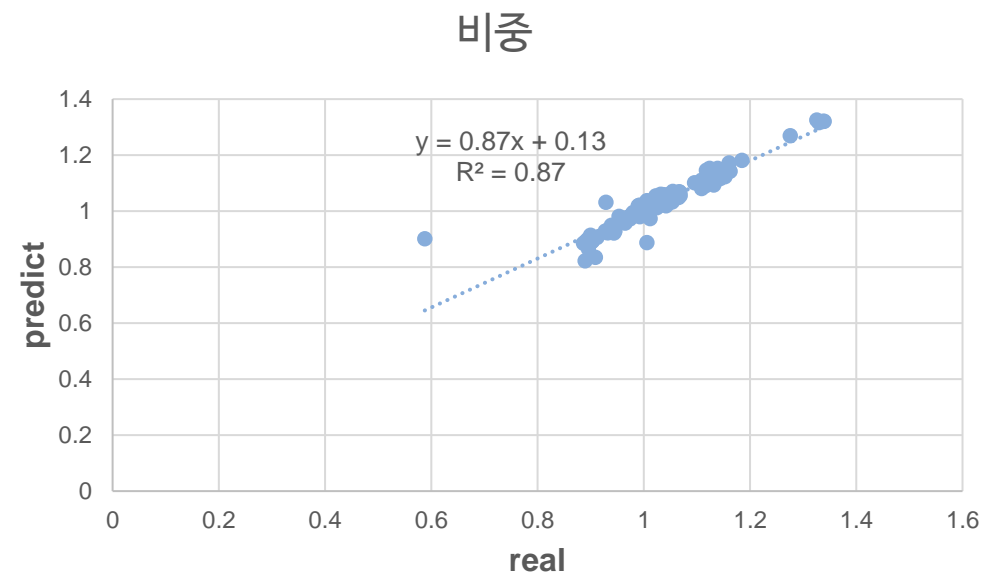
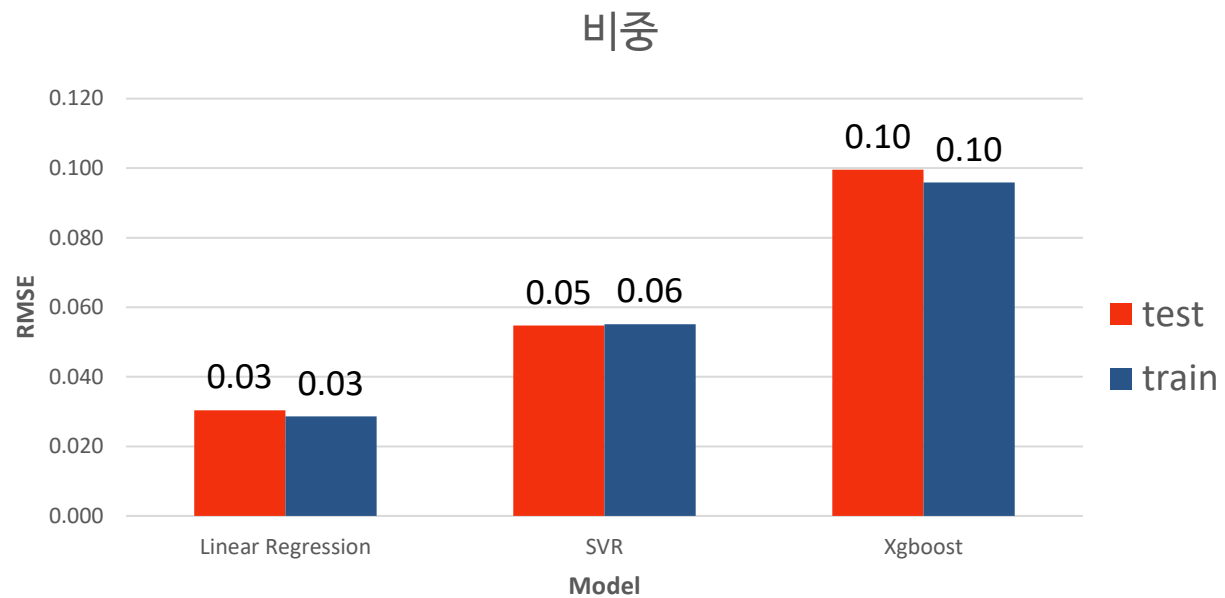
HDT, 굴곡강도,  
MI, 인장강도

MI

IZOD  
충격강도

# 03. Algorithm 선택 - Linear Regression

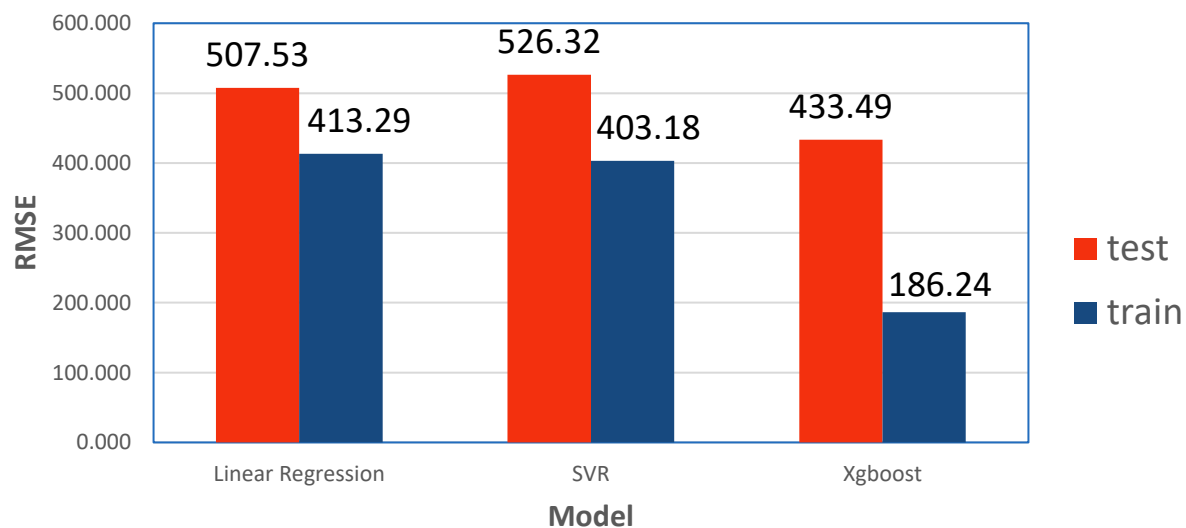
3 알고리즘



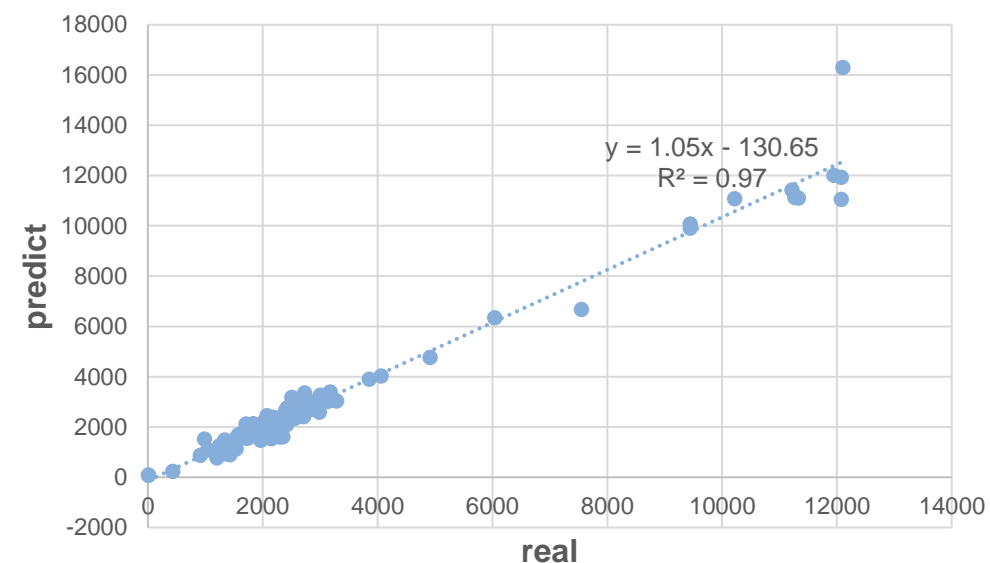
# 03. Algorithm 선택 - Xgboost

3 알고리즘

굴곡탄성률



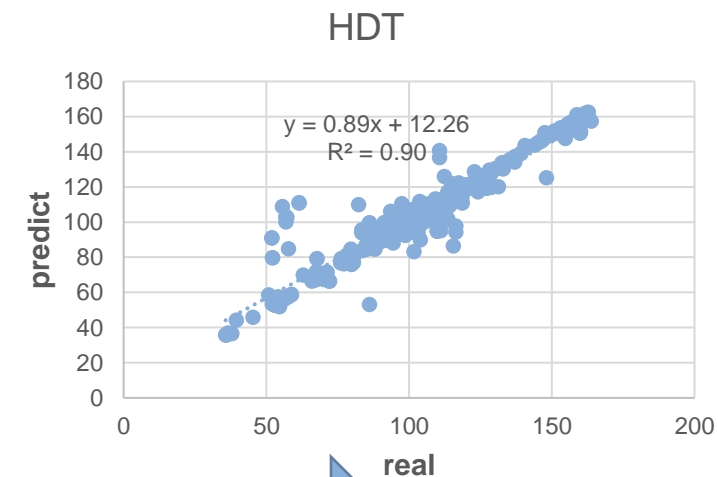
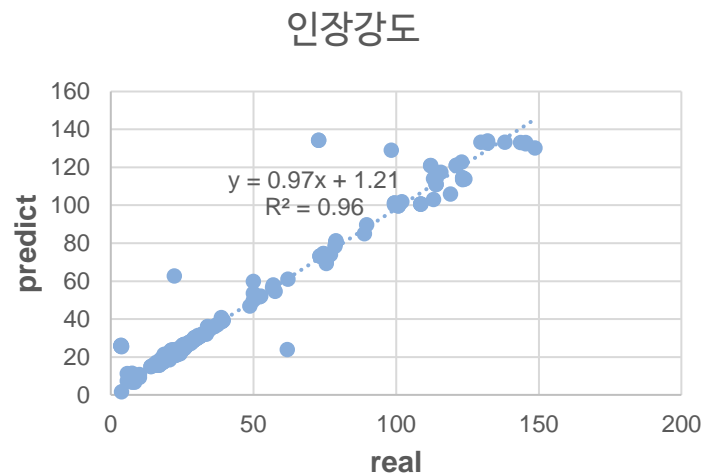
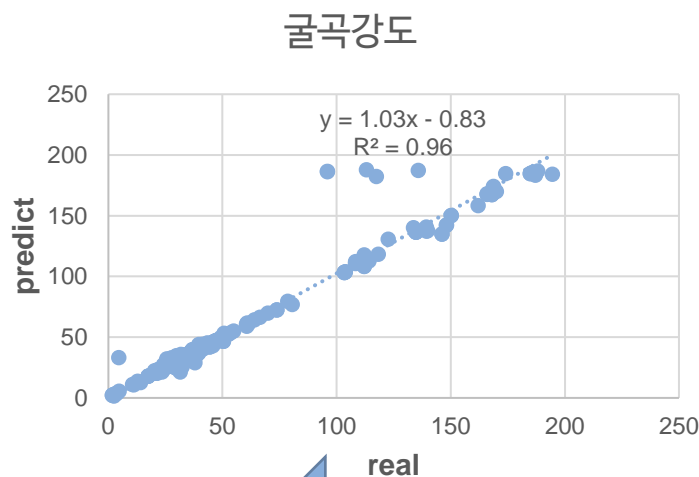
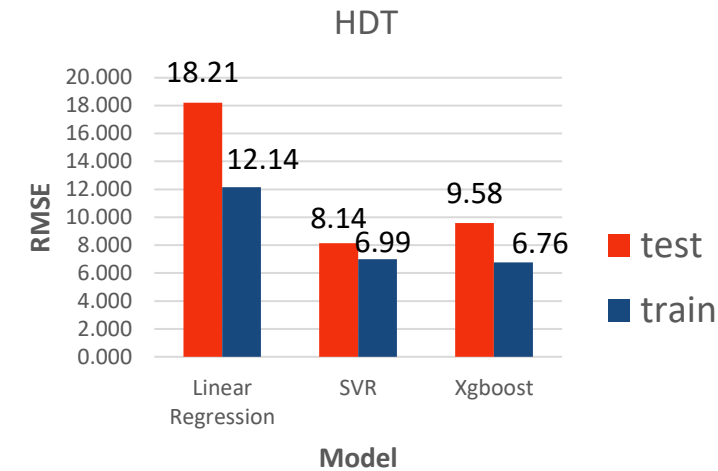
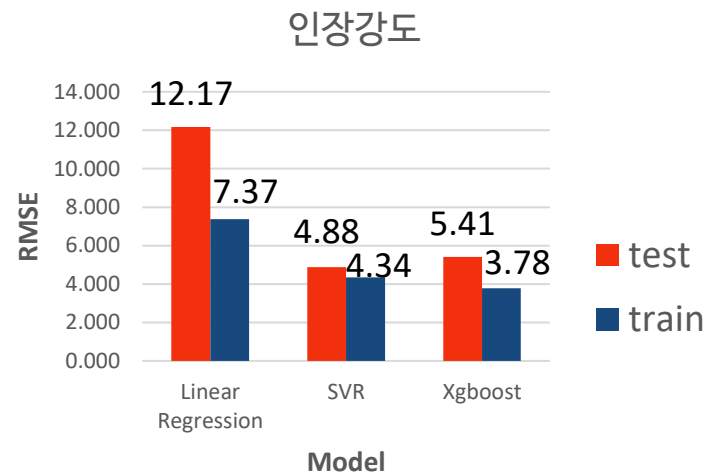
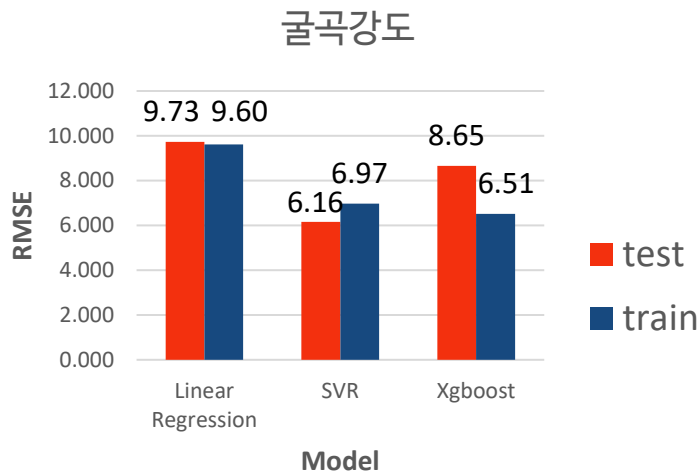
굴곡탄성률





# 03. Algorithm 선택 - SVR

3 알고리즘



비중

굴곡 탄성률

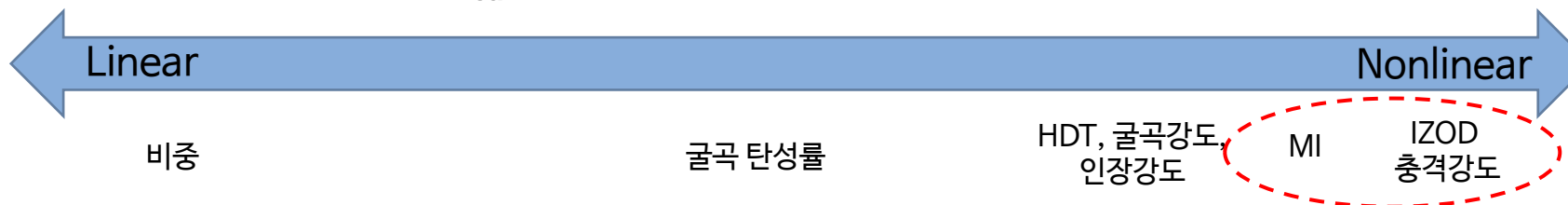
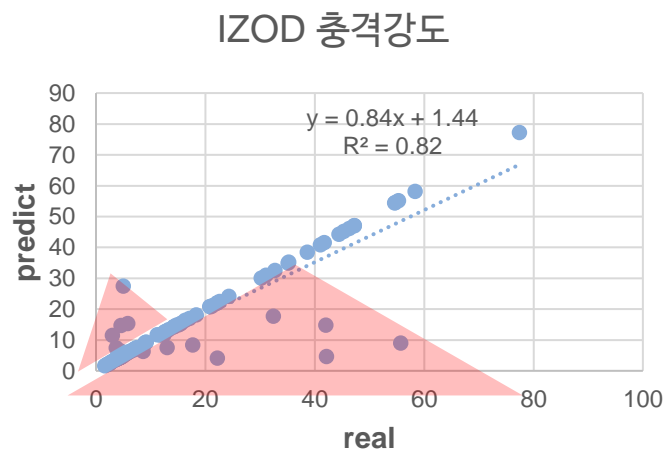
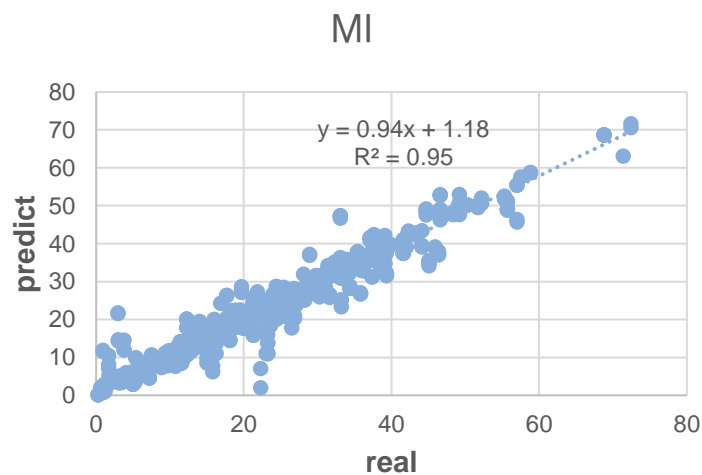
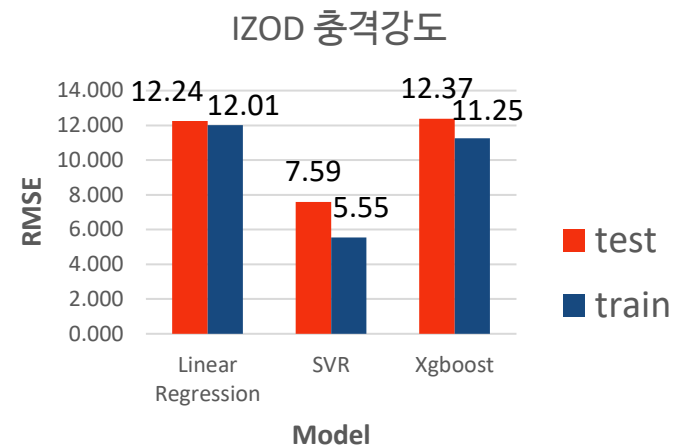
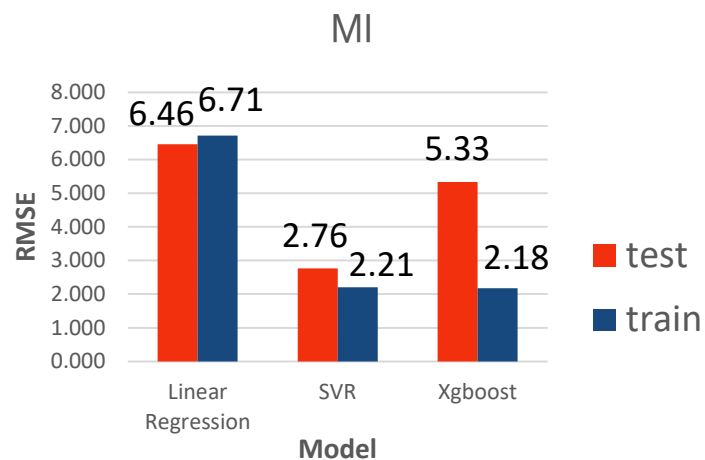
HDT, 굴곡강도, 인장강도

MI

IZOD 충격강도

# 03. Algorithm 선택 - SVR

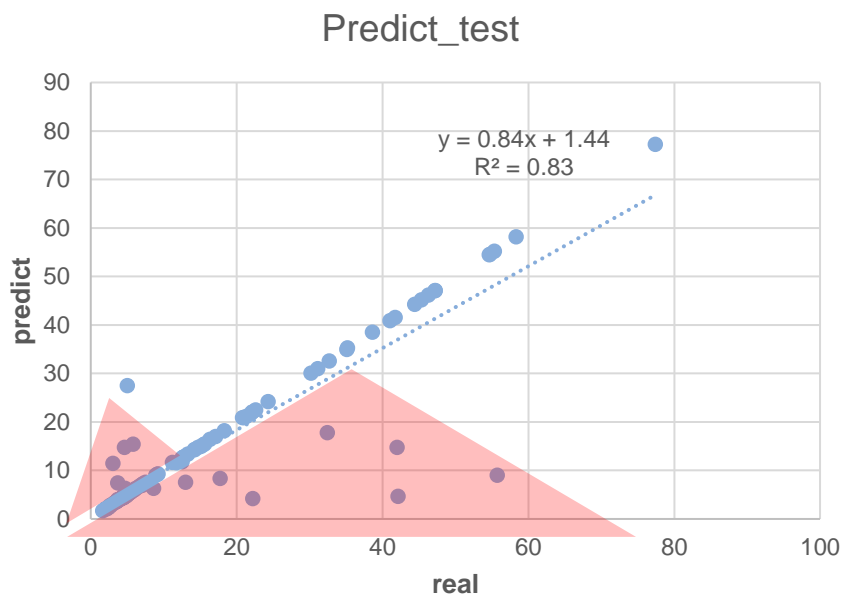
3 알고리즘





## 4. 추가 서비스

# 01. Fault Detection Alarming



## 예측 성능을 떨어뜨리는 원인?

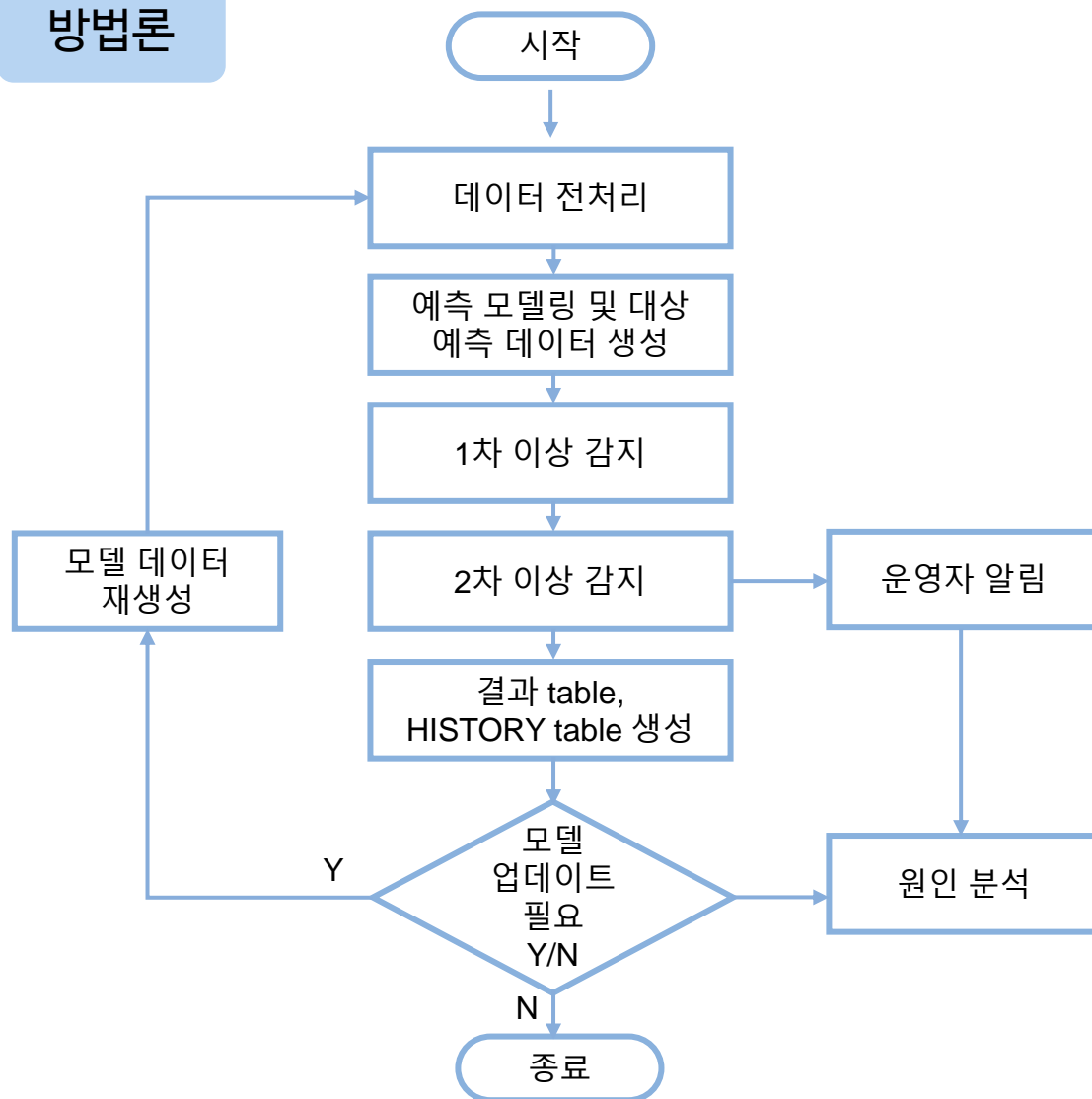
### 1. Algorithm의 문제

### 2. 데이터의 문제

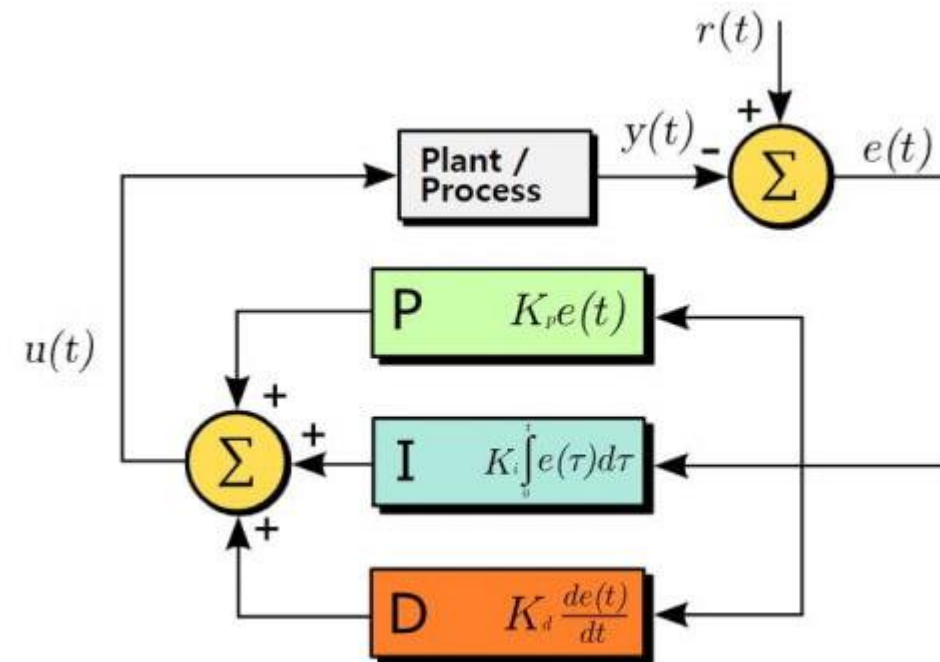
- ✓ 실험 과정에서 측정 오차
- ✓ 원재료 물성치 값 오류
  - => 기업입장에서 효율적인 연구 및 실험을 위해 엄밀한 데이터 관리가 이루어지지 못함.
  - => 공급사로부터 받은 원재료의 물성치를 실험을 통해 확인하지 않고 사용.

# 01. Fault Detection Alarming

## 방법론



## 세부 알고리즘



기준 오차지표 생성

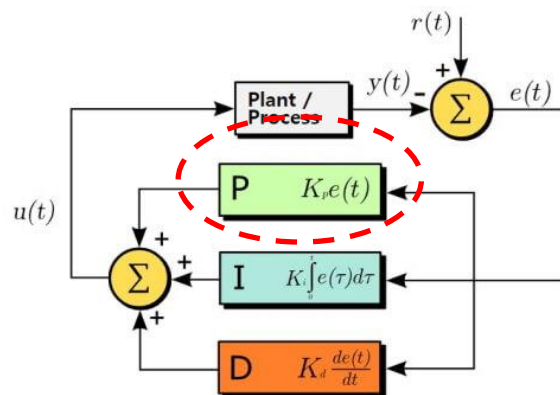
→ 종속 변수별 잔차 ⇒ SCALING ⇒ 평균

# 01. Fault Detection Alarming

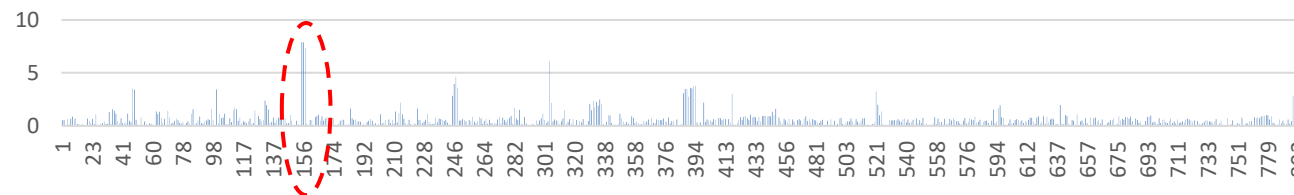
## 1차 이상 감지

- 실험에 대한 alarming

P -> 실험에서 일정 이상의 오차가 발생하는 경우 관찰



실험에 대한 detecting



155~157 실험 결과  
비중: 0.588~0.614  $\ll 0.9$



기존에 고려하지 않은 발포 공정이  
이 실험에만 사용되어 통일성을  
깨는 데이터



일시적으로 큰 오차를 갖는 실험에 대해  
alarming을 해주어 **측정 오차** 혹은 모델이  
**학습하지 못한 변수**가 발생하는 경우의  
가능성을 알려준다.



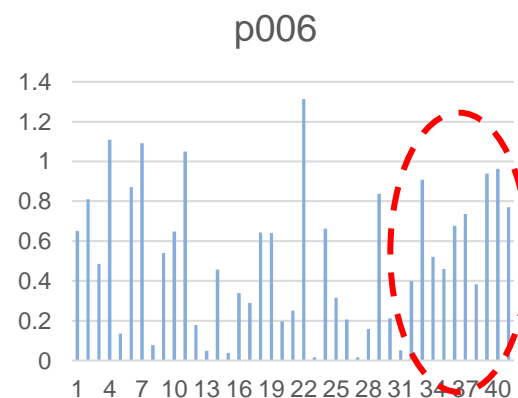
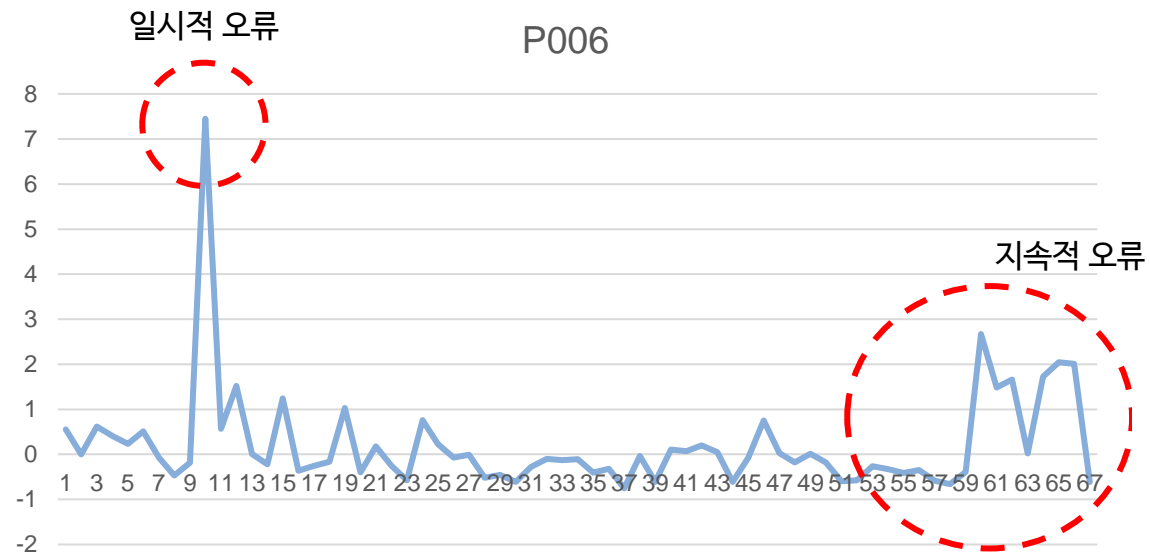
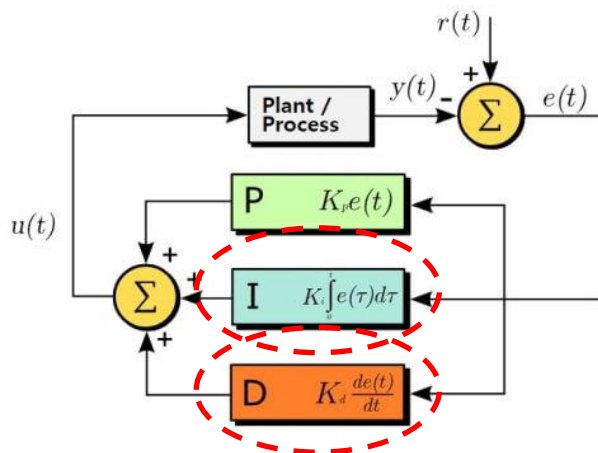
# 01. Fault Detection Alarming

## 2차 이상 감지

-재료에 대한 alarming

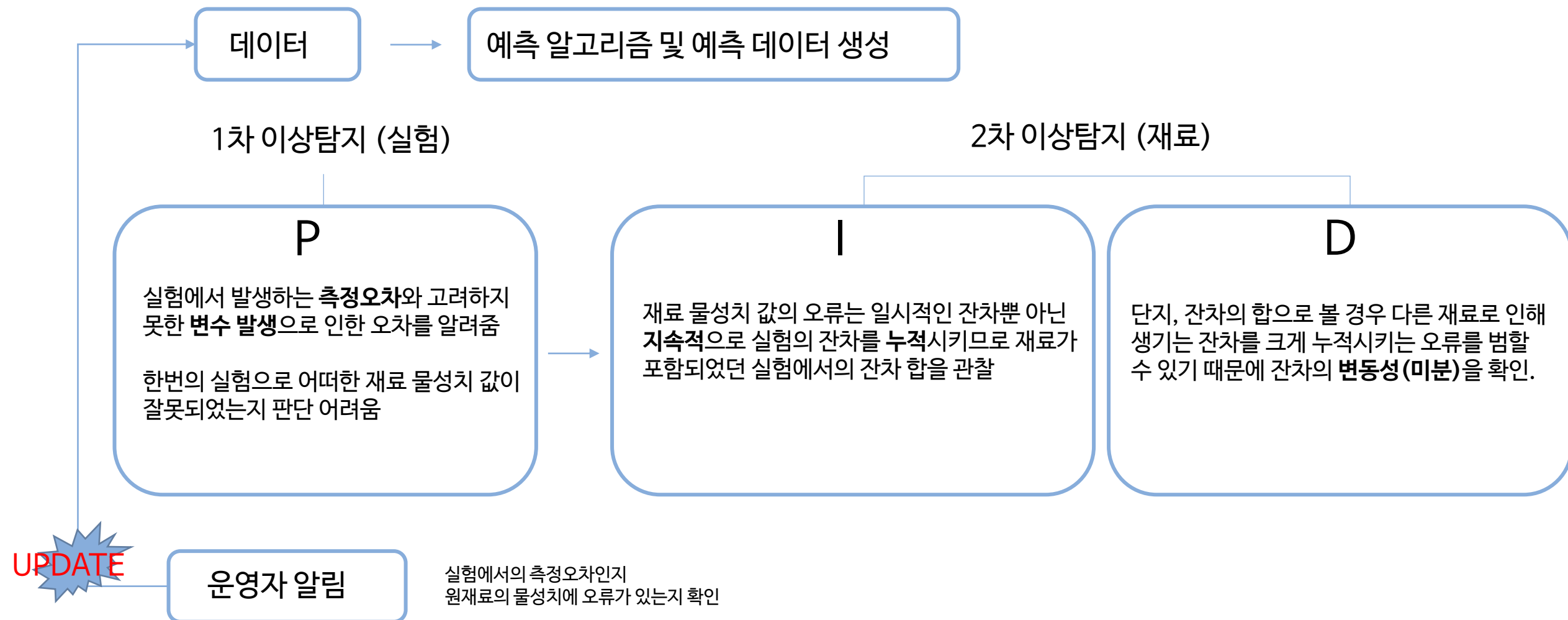
I -> 재료 단위로 실험에서 발생시킨 오차의 누적 값을 관찰

D -> 원재료 물성치 오류 Detecting



TEST 데이터이기 때문에  
생기는 오차 GAP일 수 있음.

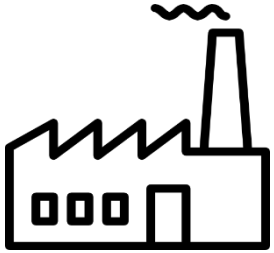
# 01. Fault Detection Alarming



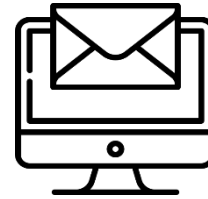
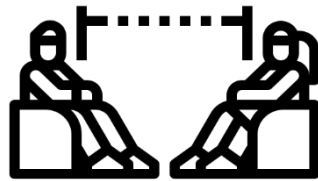
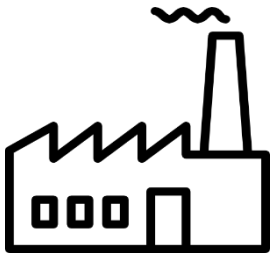
# 02. 웹 서비스

4 추가 서비스

재료 공급 업체



...



재료 공급 업체와  
고객사의 커뮤니케이션을  
도와주는 플랫폼!

→ 웹 시연

고객사

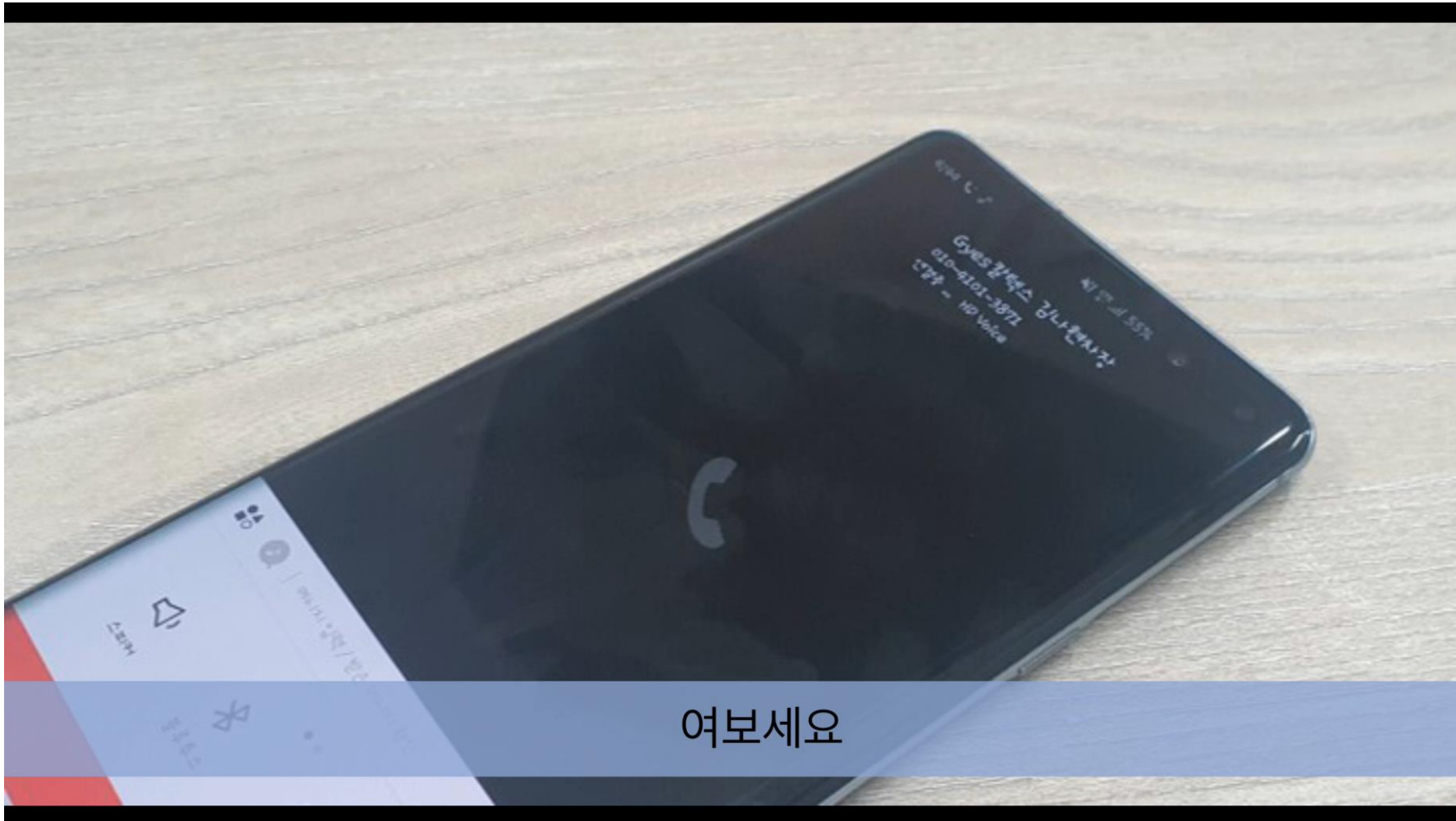


...



# 02. 웹 서비스

1 주제 선정 및 배경



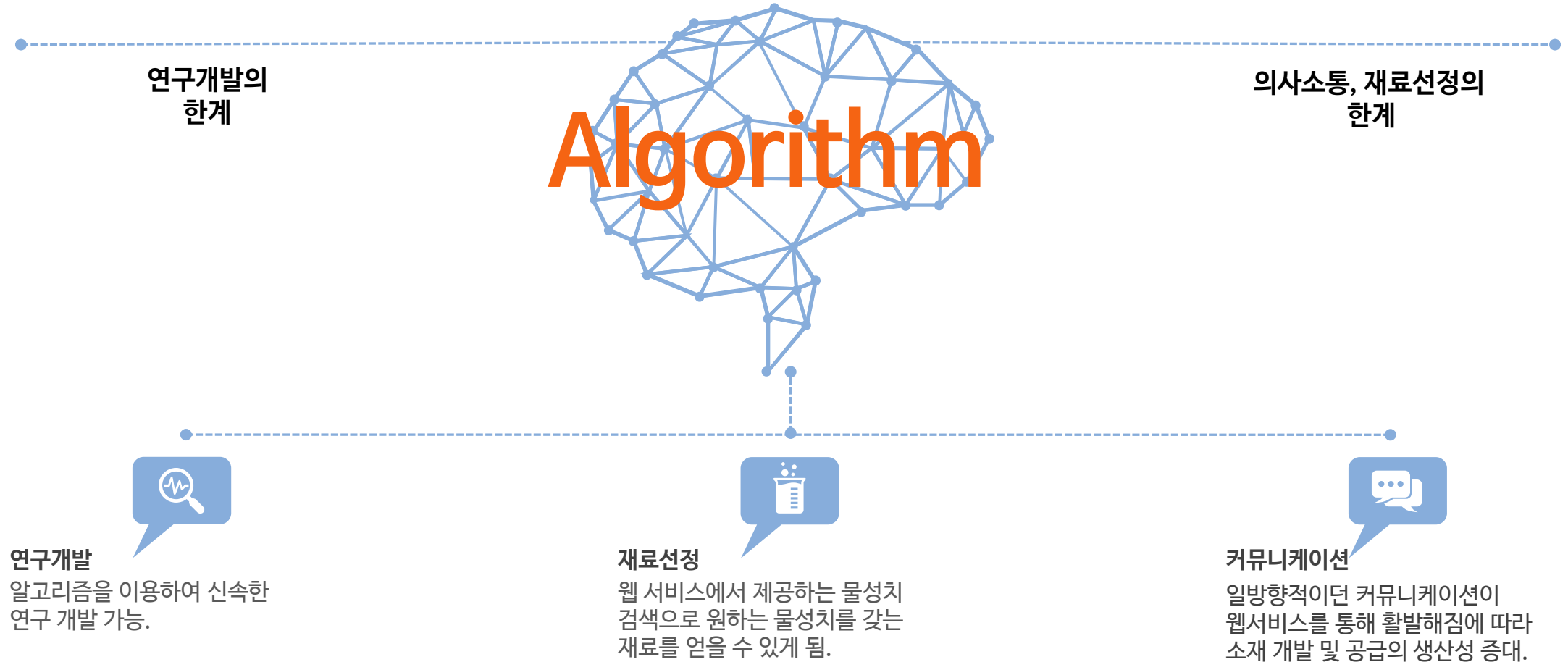
여보세요



## 5. 기대효과

# 01. 웹 서비스

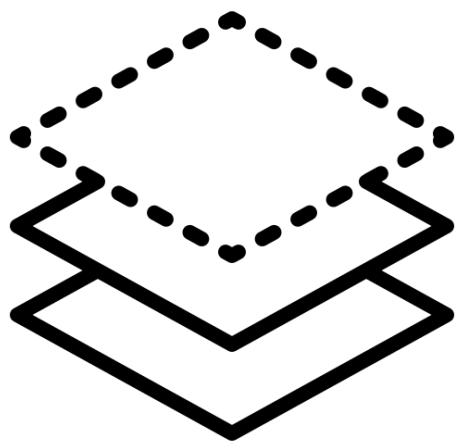
5 기대효과





# 02. Customizing

5 기대효과



Data 축적



Smart Factory



**Customizing**



GS칼텍스

에너지 솔루션 리더

A white silhouette of a human head in profile, facing right, centered on the slide.

Thank You

I am your Energy

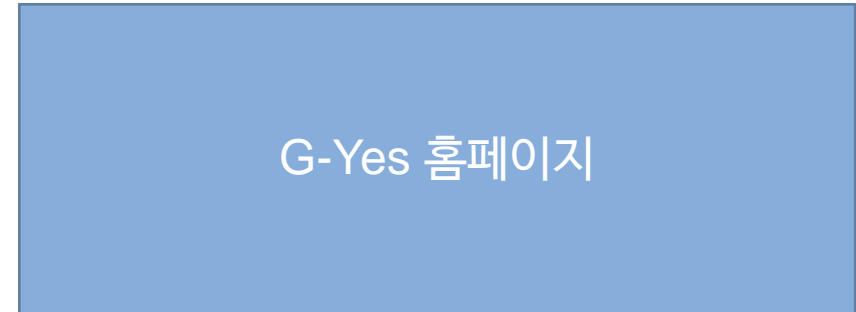
## 기존 웹 서비스와 비교



### eBooth

- 식품,플라스틱 뿐만 아니라 다양한 종류의 제품을 판매.
- 완제품을 판매.

**VS**



### G-Yes 홈페이지

- 커뮤니티를 통해 기업을 이어주는 서비스 좁은 영역의 제품을 판매 하기 때문에 고객의 니즈를 만족.
- 알고리즘을 사용해 고객의 니즈를 만족.