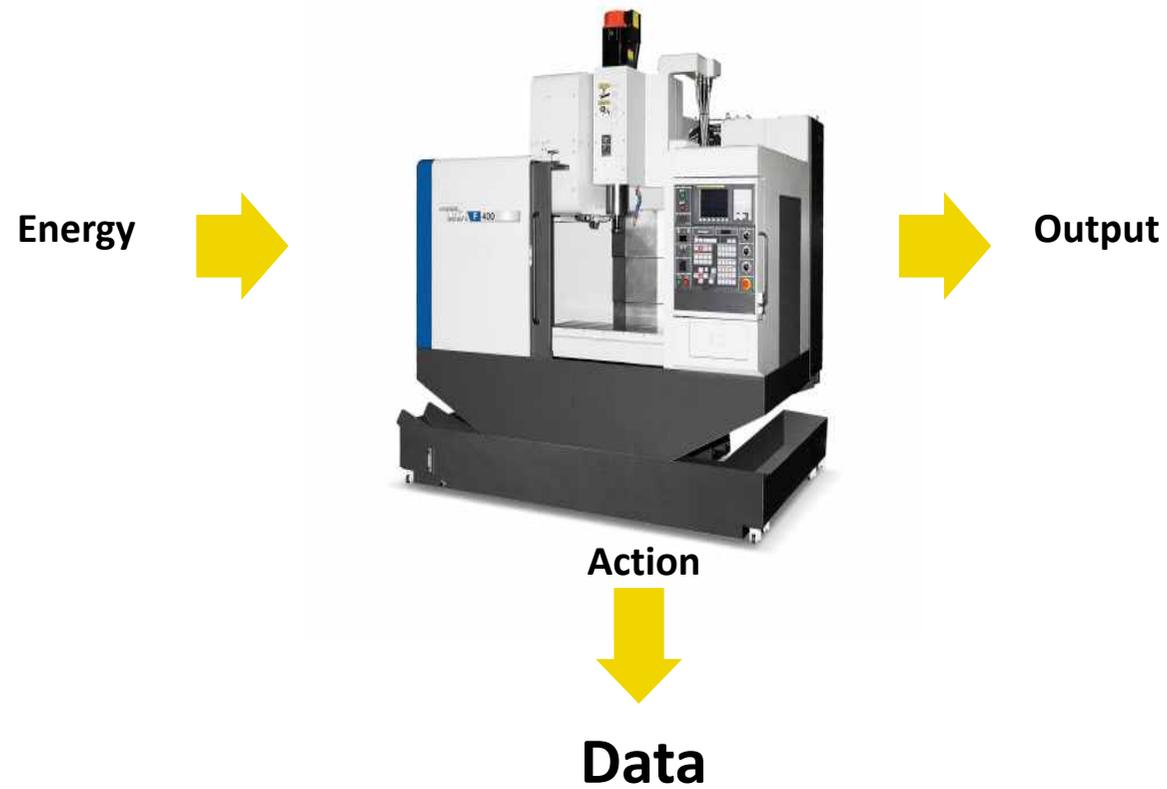


데이터 분석가를 꿈꾸는 사람들을 위한  
'데이터 분석가 취/창업의 이해와 필요역량'

임대근

# 데이터란 무엇인가?

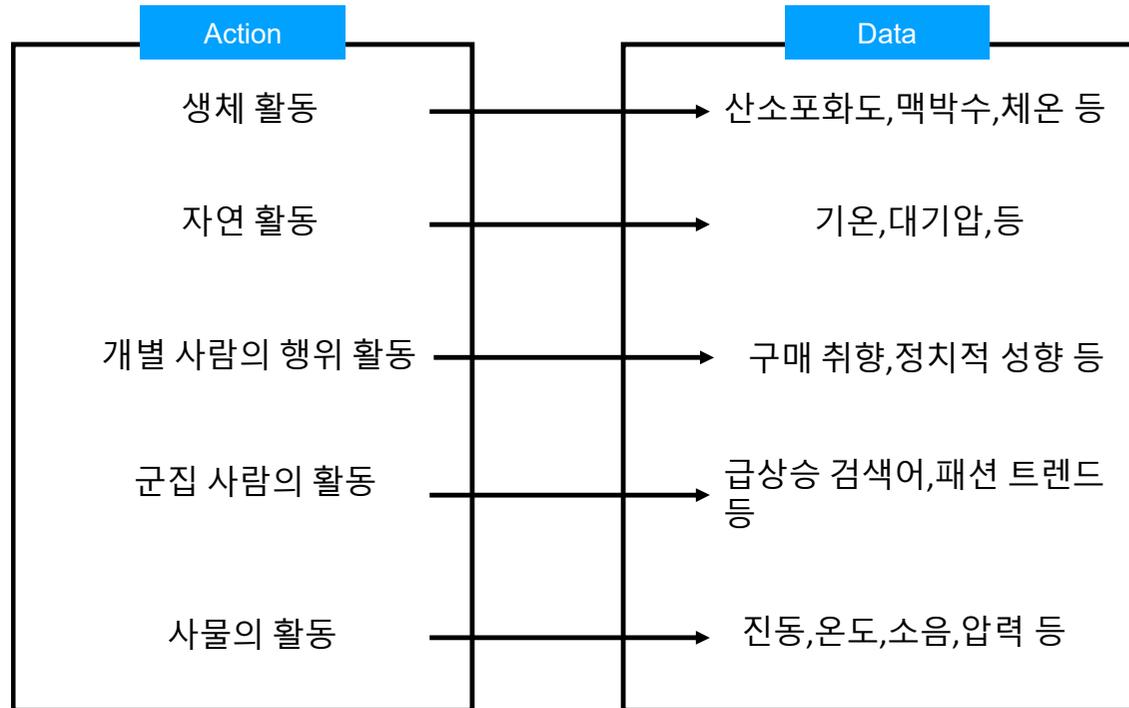


Energy를 통한 Action 의 부산물이 Data

# 데이터란 기원

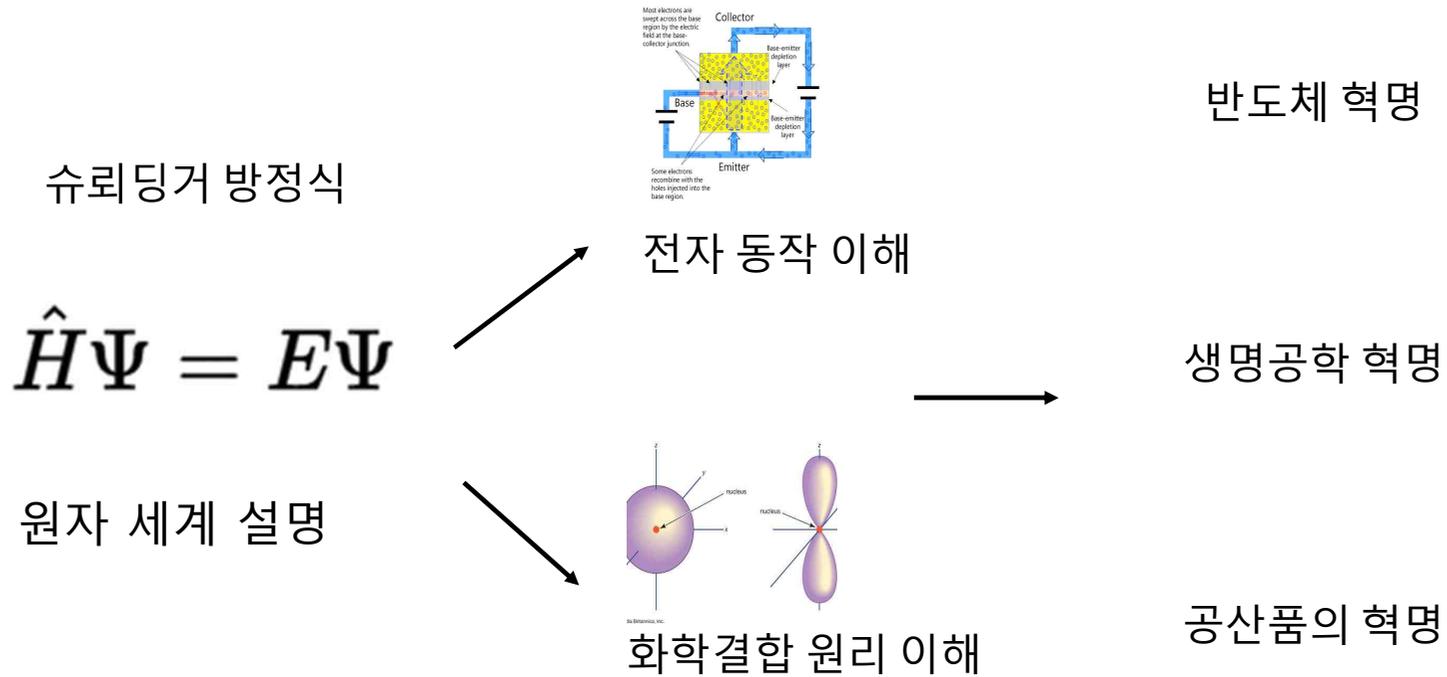


군생만물



변화 속에서 행위가 형성되고 그것의 지표가 데이터

# 현상의 숨은 원리



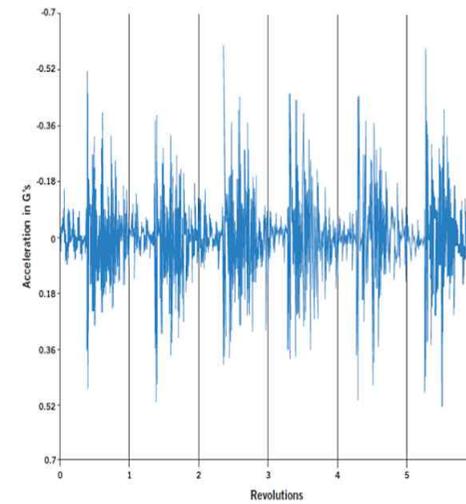
복잡한 현상 및 결과들 속에서 핵심을  
관통했던 원리는 **생각보다** 간략하고 단순 할수 있음

# 닫힌계와 열린계



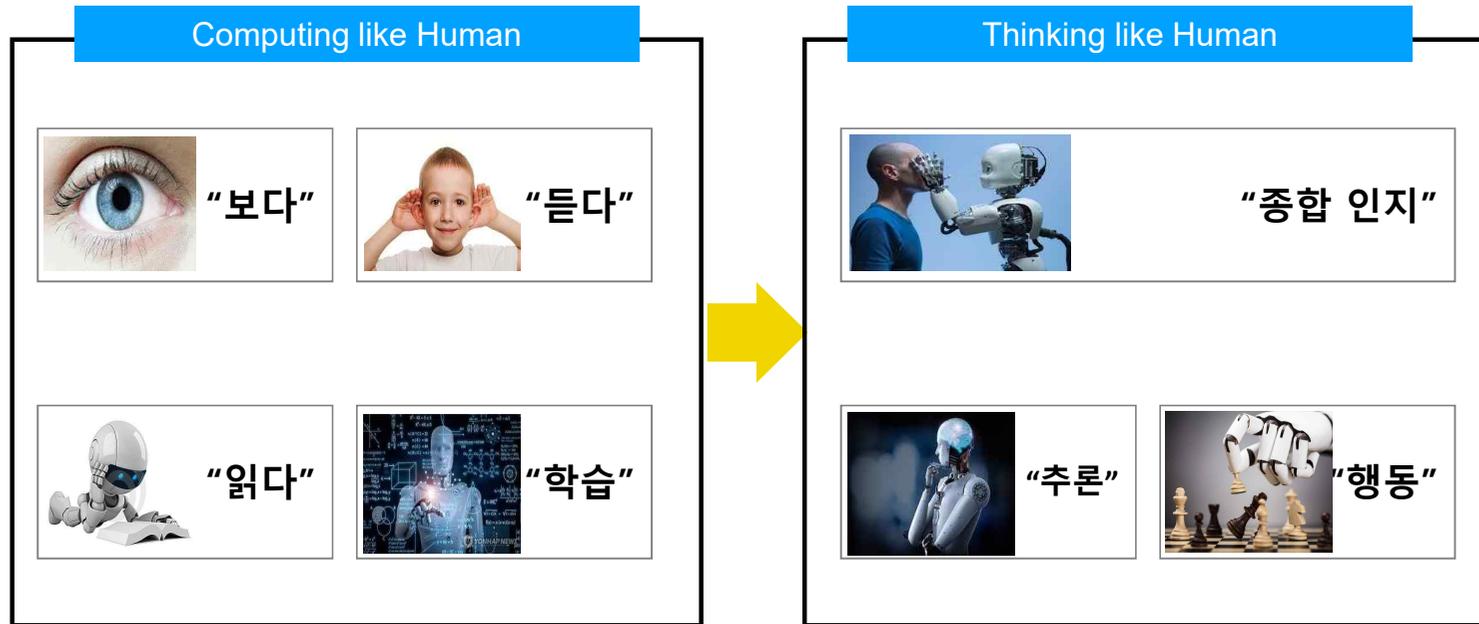
주식 데이터

VS



진동 데이터

# AI의 변화



인지 기반에서 추론으로 진화

# AI의 구분

산업용 AI는 예지 기반의 고부가가치 창출이 가능한 하이테크로 인식 위주의 마케팅 AI 와 차이가 있음

## 마케팅 AI

### 인식 위주의 AI

- 빅 플레이어(M.A.G.A)가 있음
  - Microsoft, Amazon, Google, Apple
- End User가 사람으로 성장의 한계가 있음
- AI 반도체에 이식 될 경우 서버용 위주 적용으로 성장 한계가 있음

## 산업용 AI

### 예지 위주의 AI (고부가/하이테크)

- 빅 플레이어가 없음
- End User가 사물이라 기하급수적 적용 가능
- AI 반도체의 단말단에 적용으로 활용 및 성장 가능성이 매우 높음



음성 AI



텍스트 AI



영상

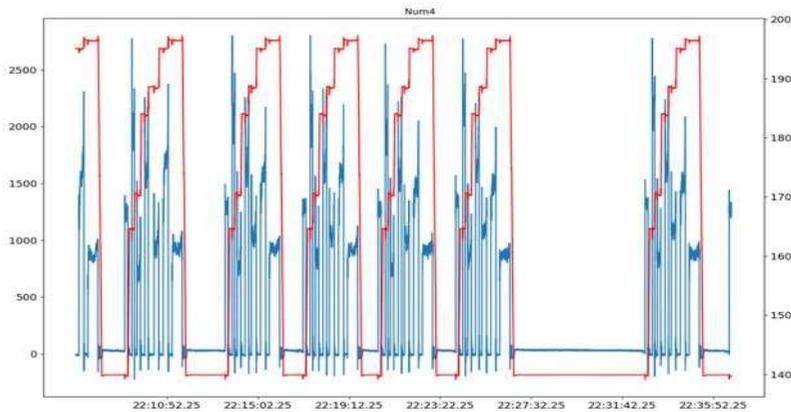


시계열 데이터 AI



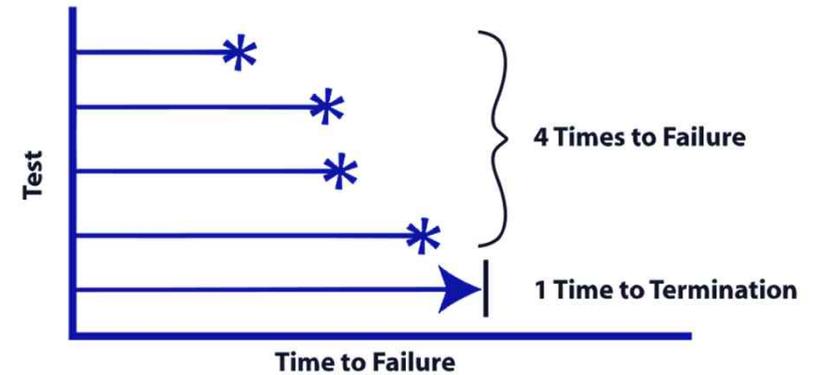
## 데이터의 실질적 문제(2/3)

산업용 데이터는 시간에 따라서  
값들이 빠르게 변하기 때문에 분석하기가 어려움



기존의 AI 방법론은 Event Data 를 통해서  
다양한 기계학습(Machine Learning) 기법을 적용했지만  
산업용 데이터는 수 ms 로 변화하는 시계열 에서  
패턴을 찾는 것이 핵심

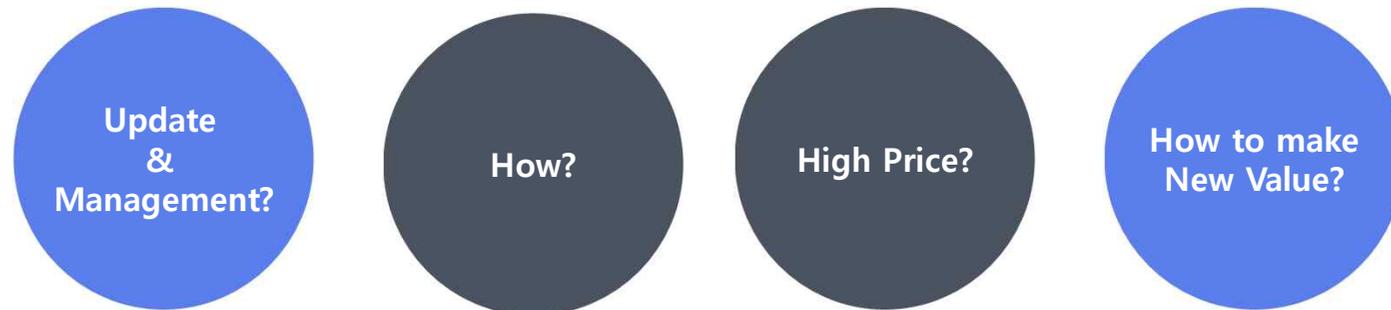
장비, 시설물에서 수집되는 데이터는 수명이 끝나기 전에  
교체하기 때문에 대부분의 데이터는 Censored Data



기존의 AI 방법론은 대부분  
Complete Data 를 사용하기 때문에  
Censored Data 를 사용하여 적용하기에는 한계가 있음

## 데이터의 실질적 문제(3/3)

지능형 서비스의 핵심은 AI 와 IoT 의 결합  
그렇지만 기업들의 AIoT Service 추진 시 **방향 · 비용 및 운영문제 직면**



*A good rule of thumb is that you should estimate that for every \$1 you spend developing an algorithm, you must spend \$100 to deploy and support it.*

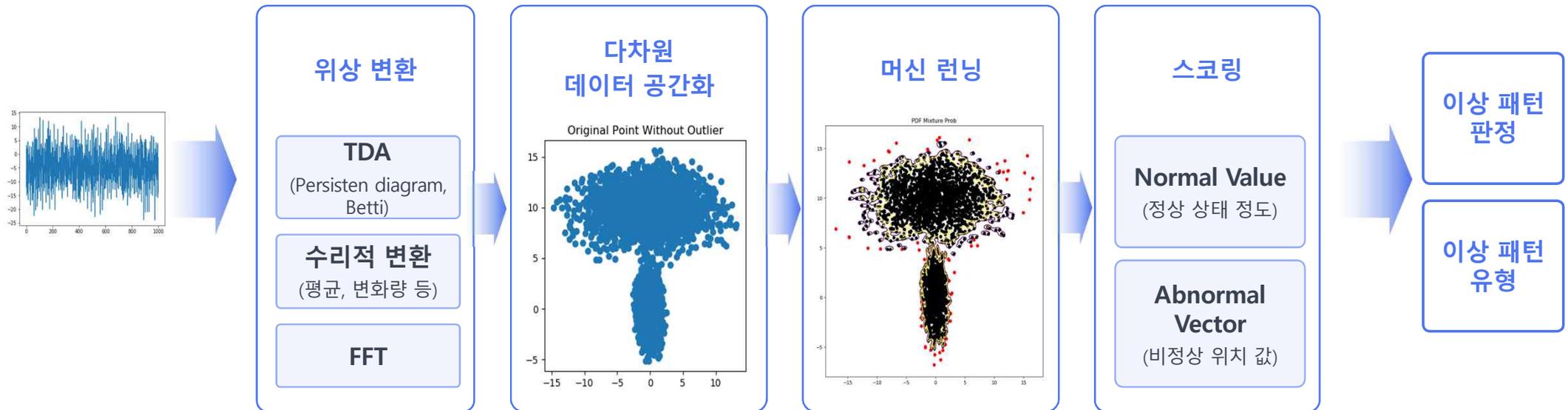
Thomas C. Redman Deloitte's Analytics and AI practice

# 시계열 데이터 분석

시계열 데이터

시계열 데이터 AI

판정 값

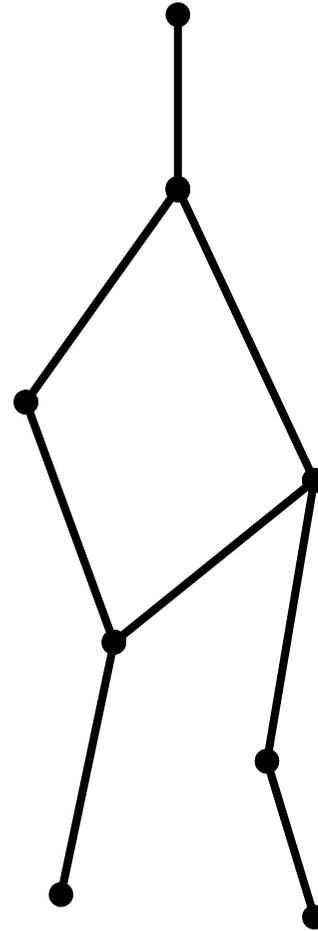


ePDF : Element Probability Density Function  
TDA : Topology Data Analysis

## 데이터의 본질화(1/3)

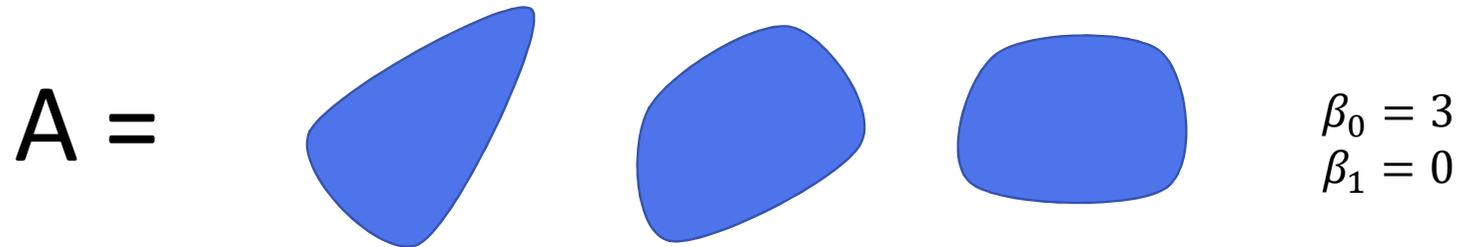


Summary



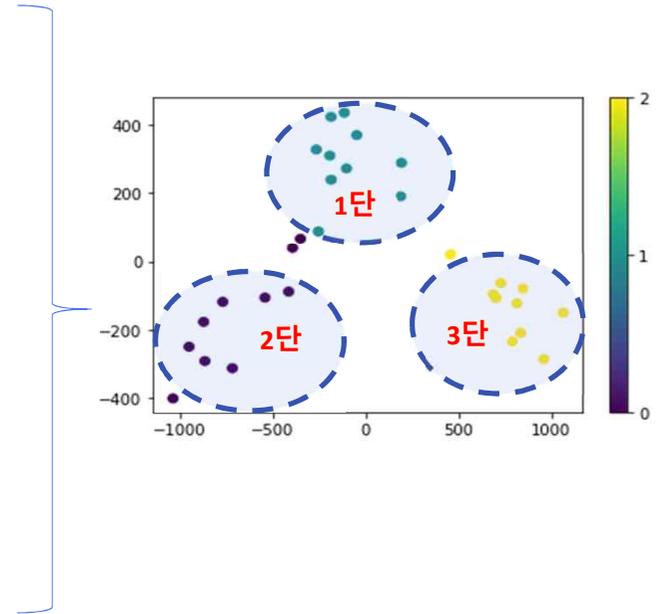
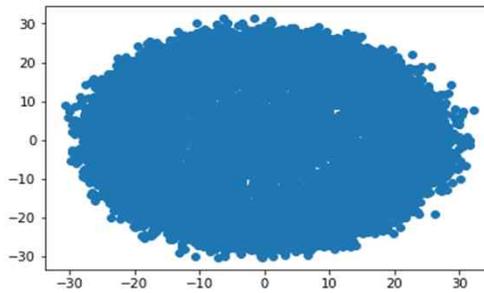
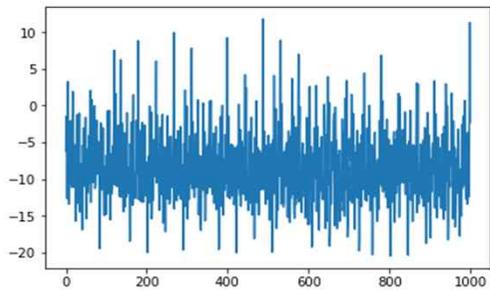
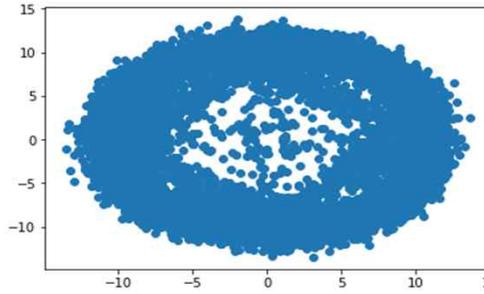
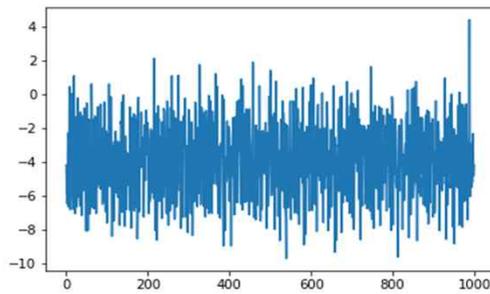
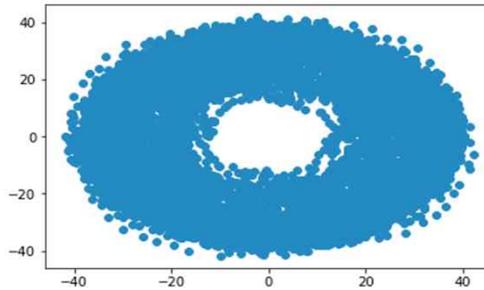
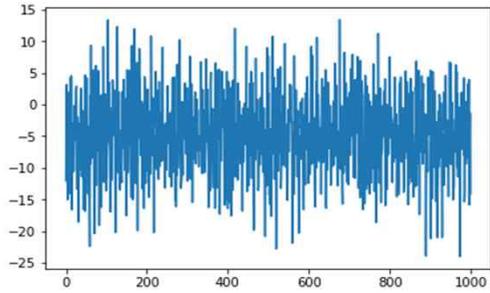
## 데이터의 본질화(2/3)

Topological space를 구별하기 위해, 개수( $H_0$ ) 및 구멍( $H_1$ ) 등을 세는 법



# 데이터의 본질화(3/3)

3단



# AI 적용 로드맵



### Light AI

- 데이터 간의 연관관계 분석을 통한 이상지점 분석
- 단순 회귀분석을 통한 피크 치 예측

**데이터 분석 위주로 보조적 AI**

### High-end AI

- 다변수 복합 분석을 통한 고장예측
- 결과 피드백을 통한 최적화 값 제시

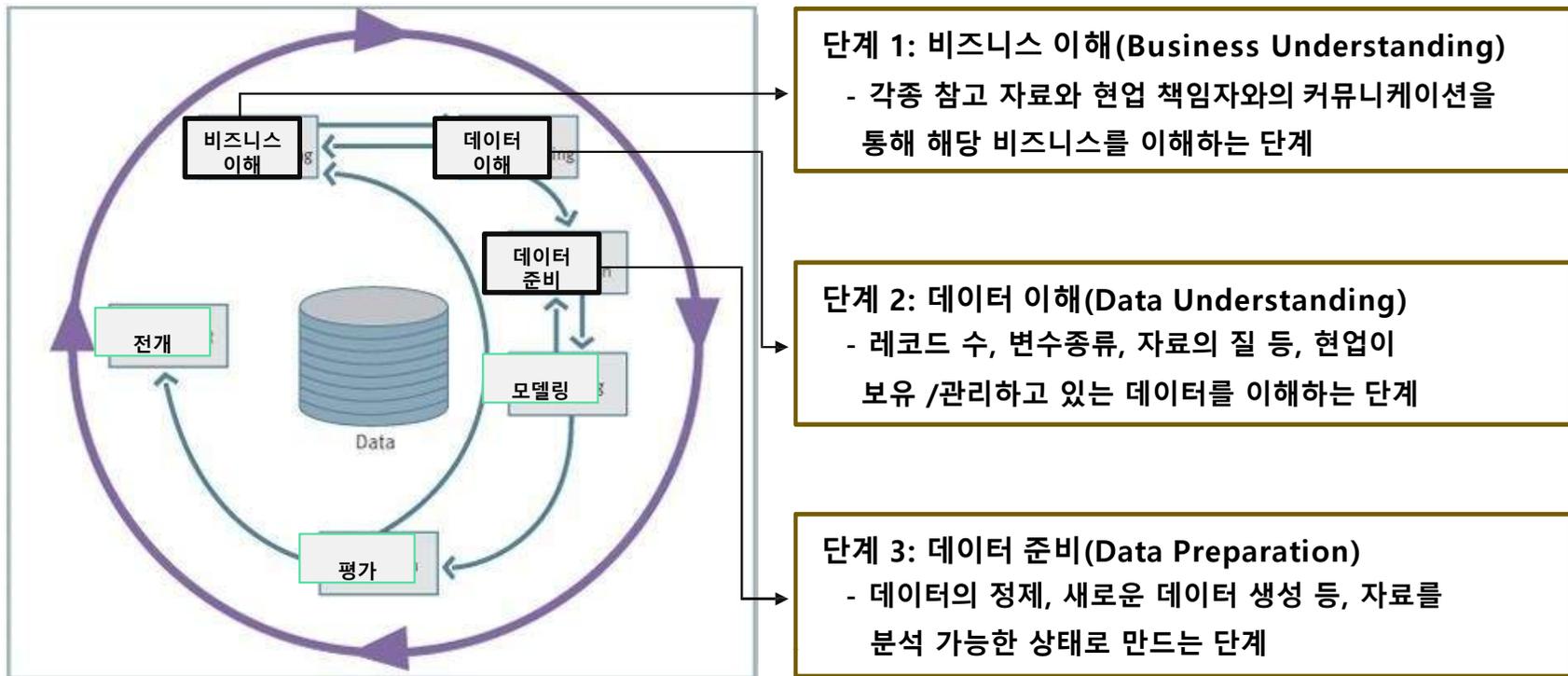
**데이터 분석 및 학습 알고리즘 적용을 통한 예지적 AI**

# B2B 서비스 관점에서 AI

Set-and-Forget Service		Set-and-Start Service
ERP/UMS/협업툴/보안/MES/Devtool	분야	IoT / AI
초기 개발 비용 및 구축시간이 필요하며 추가적인 변경 사항은 한정적으로 발생	구축 방식	구축 이후에도 꾸준한 관리 및 지속적인 업데이트 필요
특화 개발/적용 후 최소 관리	데이터 특징	서비스 적용 후 지속적 업데이트
Historical Data	가치	Realtime Data
업무 효율화를 통한 비용 절감	가성비 높은 구현 방향	업무효율화를 통한 비용 절감 가치 창출을 통한 매출 증대
PaaS 기반으로 성과 극대화 가능		SaaS 기반으로 성과 극대화 가능

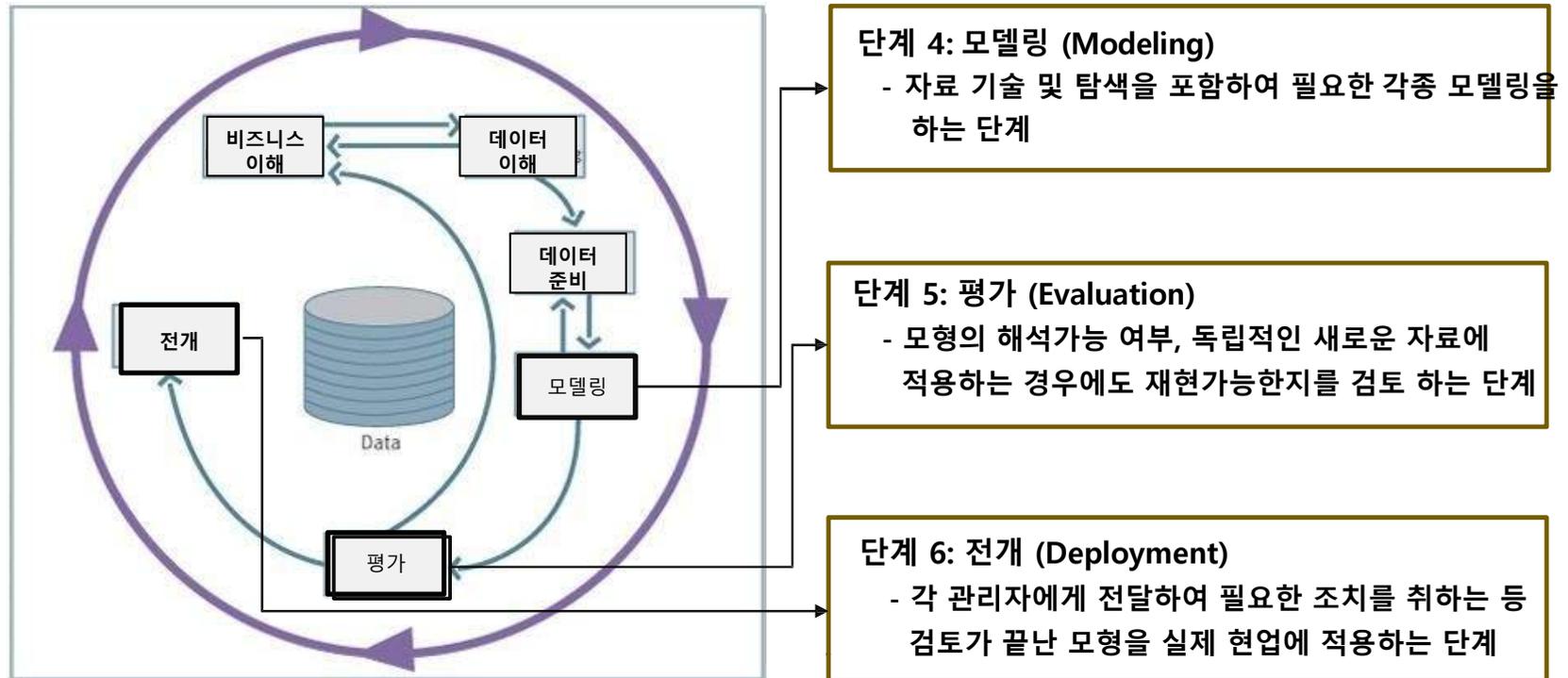
# AI 적용 과정

## 데이터 분석 모델링 프로세스 (1/3)



# AI 적용 과정

## 데이터 분석 모델링 프로세스 (2/3)



# AI 적용 과정

## 데이터 분석 모델링 프로세스

