

인공지능 (Artificial Intelligence)

Aug, 2024

훈련일정

	오전	오후
1차시 13일(화)	1교시 : Data Literacy, 빅데이터, 인공지능 개론	4교시 : 구글 코랩 활용(모딕 1장), 모딕 2장(딥러닝 핵심)
	2교시 : 인공지능의 활용 사례(등대공장)	5교시 : 파이썬 빅데이터 분석 실습1 (numpy, pandas)
	3교시 : 인공지능 개념, Cloud, 통계 기초	6교시 : 파이썬 빅데이터 분석 실습2 (pandas, matplotlib)
2차시 14일(수)	1교시 : 모딕 데이터 분석 실습(11장 피마인디언, 12장 붓꽃)	4교시 : 모딕 4장(가장 훌륭한 예측선, 최소제곱법, 평균오차)
	2교시 : 데이터 분석연습(와인, 타이타닉)	5교시 : 선형회귀 실습 - 보스턴 주택가격 예측
	3교시 : 모딕 3장 (기초 수학)	6교시 : 선형회귀 실습 - 보스턴 주택가격 예측 / 자동차 연비 예측
3차시 16일(금)	1교시 : 선형회귀 실습 - 자동차 연비 예측	4교시 : 선형회귀 실습 - 당뇨병 예측
	2교시 : 모딕 5장 (경사하강법, 선형회귀, 다항회귀)	5교시 : 모딕 7장 신경망(퍼셉트론)
	3교시 : 모딕 6장 (로지스틱회귀)	6교시 : 모딕 8장 (다층퍼셉트론)



CONTENTS

1. 데이터 문해력(Data Literacy)
2. 빅데이터
3. 인공지능(머신러닝, 딥러닝)
4. 머신 러닝 기초 및 원리
5. DX, Smart(Factory, City, Mobility, Logistics)
6. 기업 활용 사례

Q1) 가제

A1) 가재



Q2) 사흘

A2) 3일간이란 의미 모름

Q3) 금일

A3) 금요일

Q4) 고지식

A4) 높은(高) 지식

문맹률? 문해율?
차이를 아시나요?

바늘링코



문맹률?

글을 배우지 못해서 읽거나
쓰지 못하는 사람의 비율

바늘링코



출처 : 『EBS 미래교육 플러스 - 배움의 기초, 문해력』 -2020년

우리나라의 경우 기본 문맹률 1%로
글을 못읽는 사람이 거의 없습니다.

바늘링크



그렇다면 글을 읽을 줄 안다고
문장을 이해하고 있을까요?

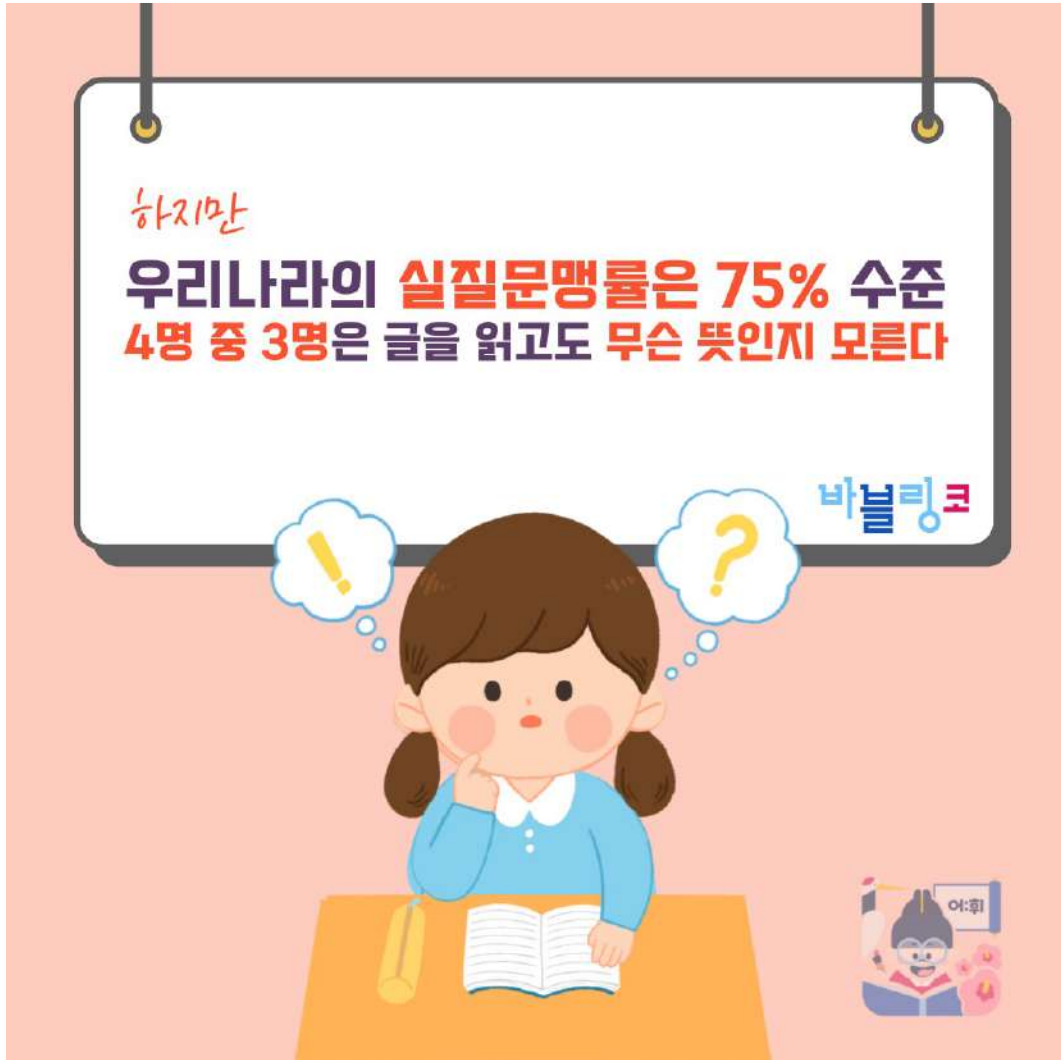
바늘링크



출처 : 『EBS 미래교육 플러스 - 배움의 기초, 문해력』 -2020년



출처 : 『EBS 미래교육 플러스 - 배움의 기초, 문해력』 -2020년



문해력(Literacy)

정보를 이해하고 받아들이는 능력

문해력 키우기

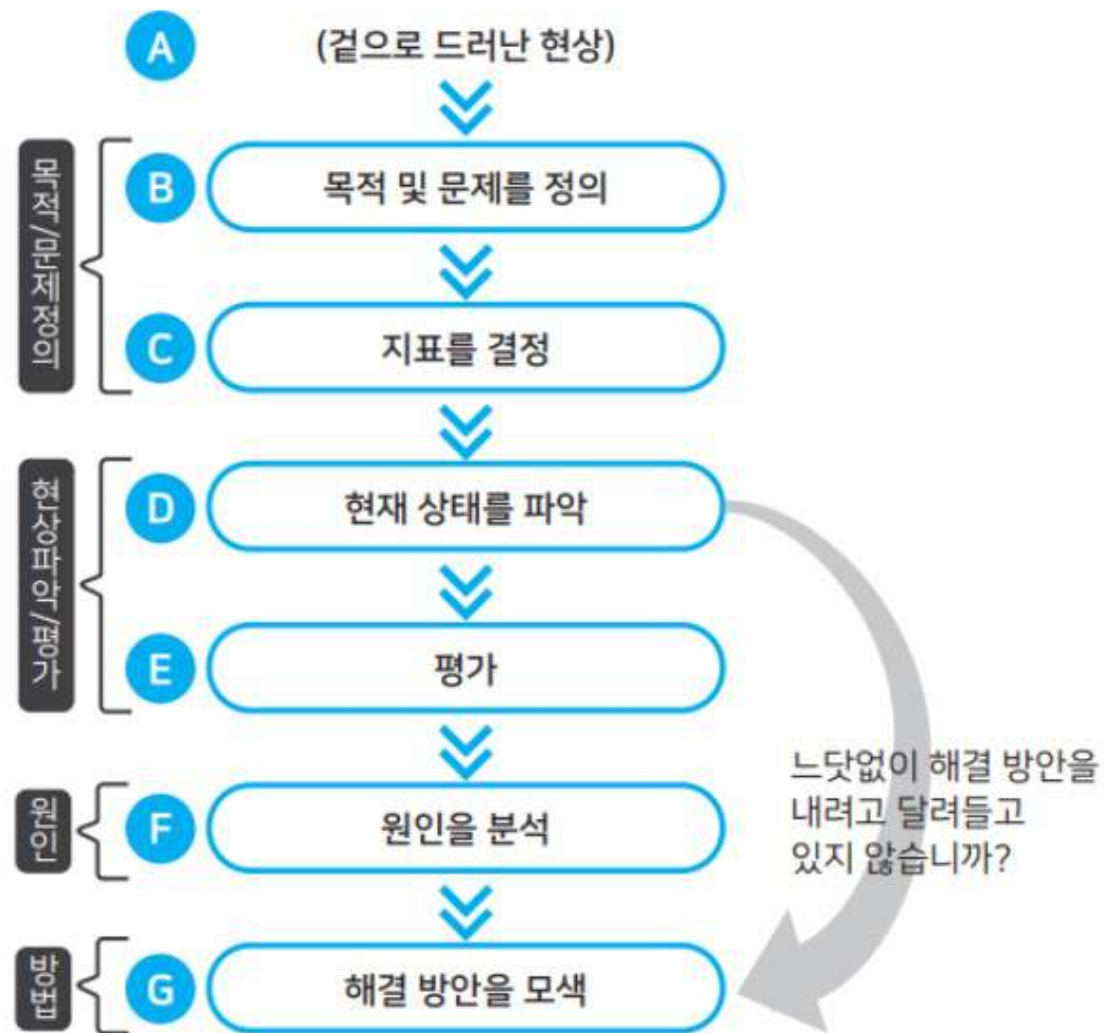
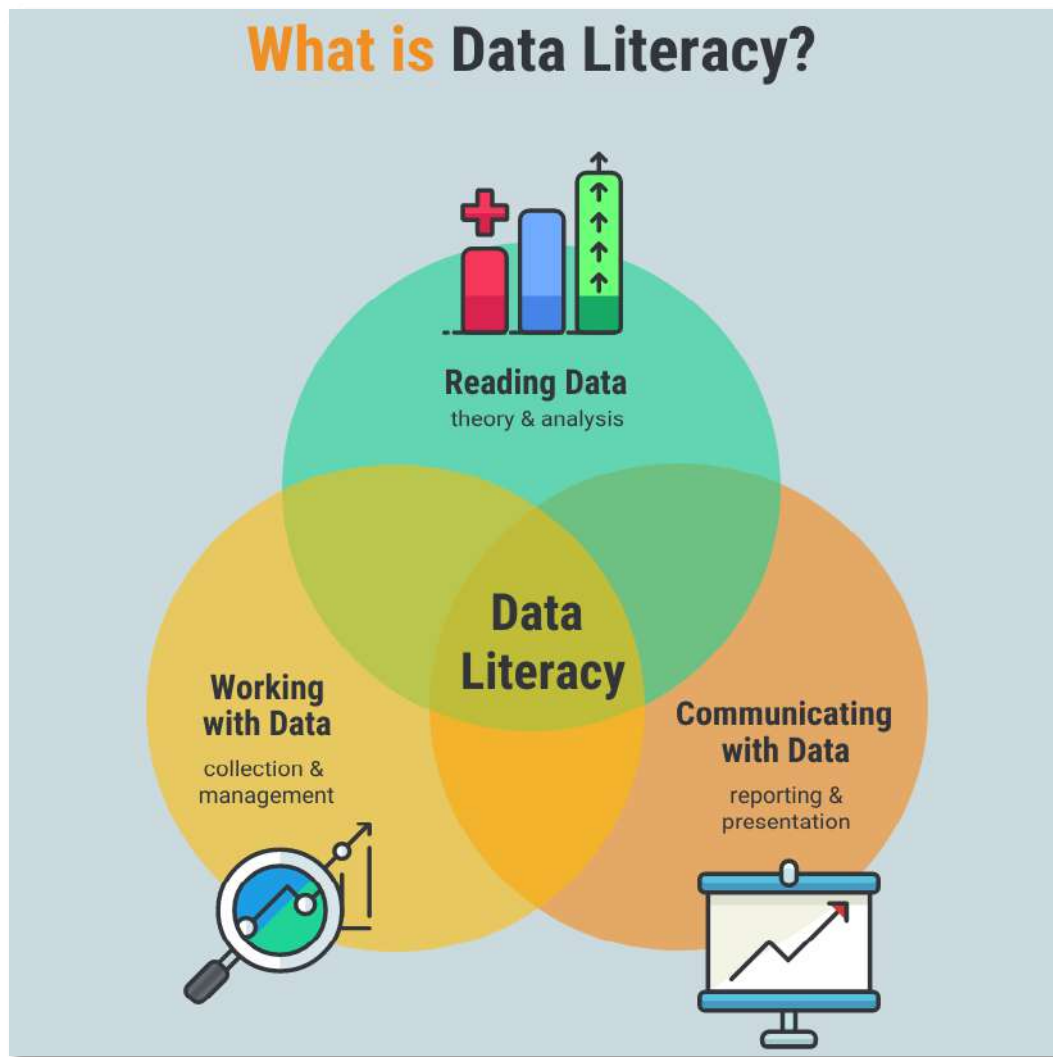
글을 읽고 글 사이 관계 파악하기

질문하는 힘 기르기



출처 : 『EBS 미래교육 플러스 - 배움의 기초, 문해력』 -2020년

데이터 문해력(Data Literacy)





빅데이터 출현 배경

정보기술의 발전에 따른 컴퓨터 활용의 확대

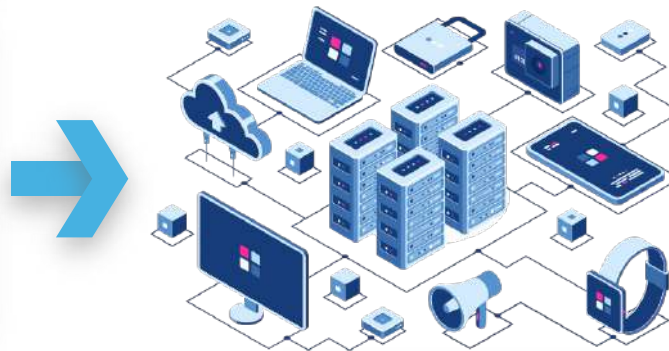
소셜 미디어(Social Media)의 급격한 확산

IoT 확산에 따른 센서 데이터의 확대

멀티미디어 콘텐츠와 사용에 대한 정보 급증

기업의 고객 데이터 트래킹 및 수집 행위 증가

데이터 저장 및 처리 기술의 발전



인프라와 기술

ICT 주도권의 이전



빅데이터란 무엇인가?

"BIG DATA ≠ 대용량 자료" 가 아니다.

- 조직의 내외부에 존재하는 다양한 형태의 데이터를 수집, 처리, 저장하여
- 목적에 맞게 분석함으로써 해당 분야의 필요 지식을 추출하고
- 이를 조직의 전략적 의사결정에 활용하거나
- 이를 시스템화하여 상시적으로 생산성향상에 활용하거나
- 새로운 비즈니스 모델의 창출에 활용하고자 하는 패러다임

과학적 논리 전개방식

연역

이미 알고 있는 일반적인 지식, 법칙, 원리로부터 합리적 추론에 따라 필연적 결론을 이끌어 내는 접근방법

귀납

개별 사례들에 대한 관찰을 통해 일반적인 결론을 이끌어 내는 접근방법

빅데이터는 전형적인 귀납적 연구 방법론에 해당

귀납법 전개방식을 위해서는

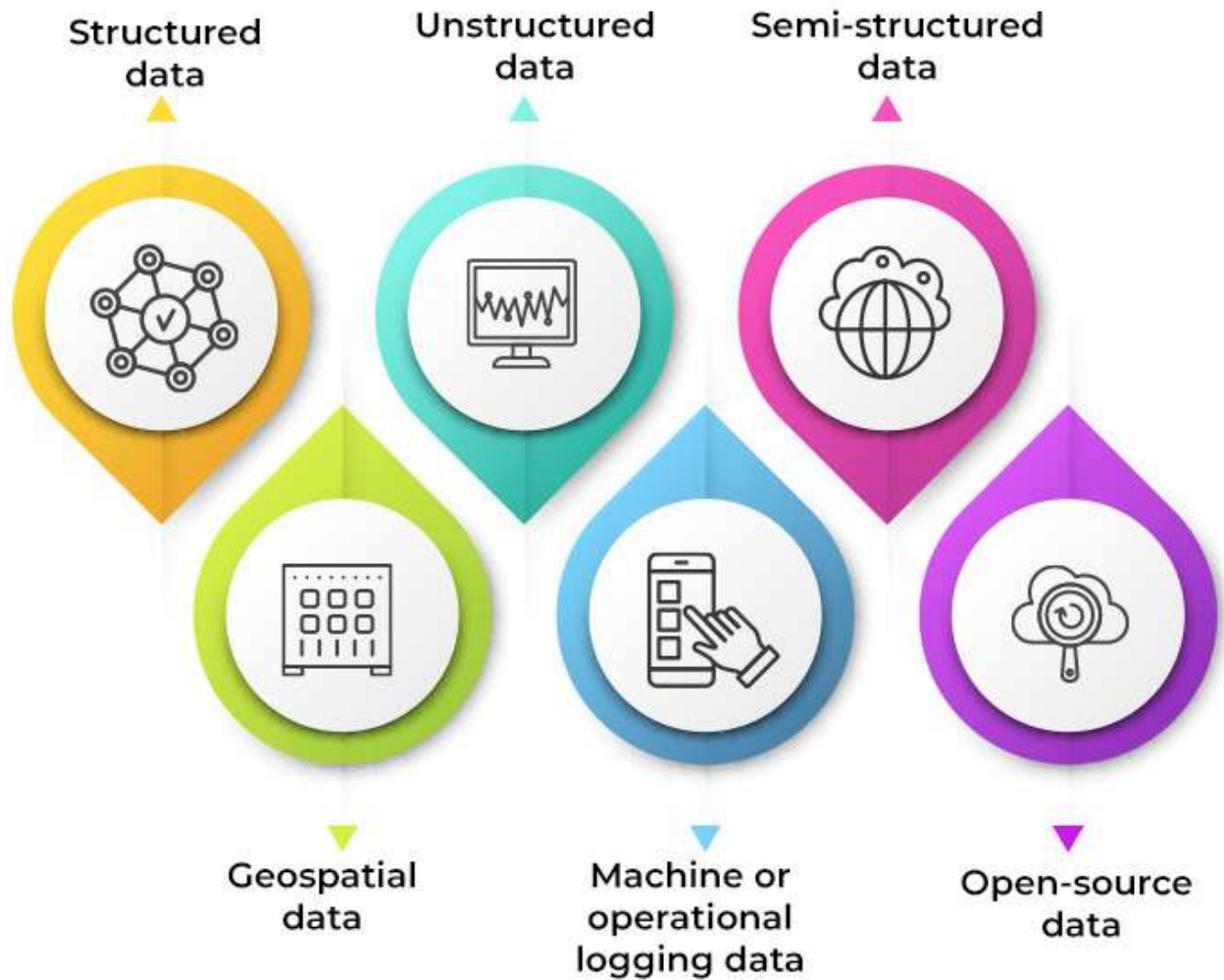
1. 데이터
2. 저장장치
3. 컴퓨팅 파워가 전제조건

Big Data의 정의



1. 3V → 4V → 5V
2. 처리, 분석 기술적 변화까지 포함
3. 인재, 조직 변화

TYPES OF BIG DATA



DATA



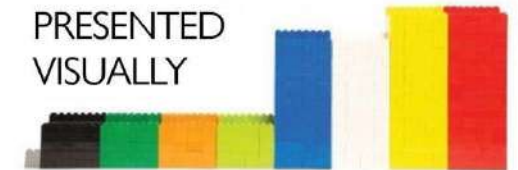
SORTED



ARRANGED



PRESENTED VISUALLY



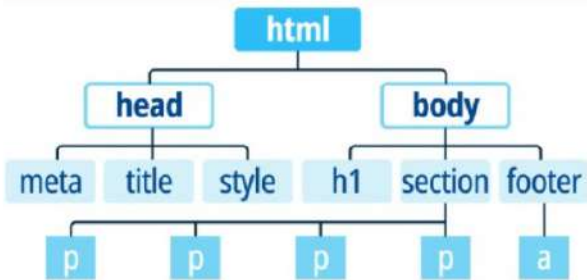
EXPLAINED WITH A STORY



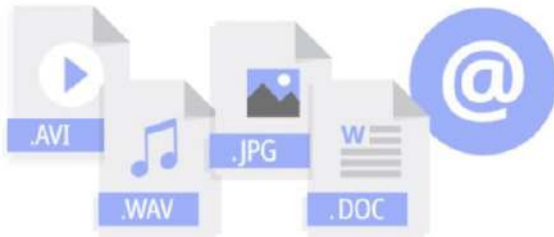
Big Data

ID	Name	AGE	SEX
01	KIM	32	M
02	LEE	26	F
03	PARK	72	F
04	CHOI	15	M

structured
data



semi-
structured
data



unstructured
data

정형데이터

1

- RDB
- Spreadsheet
- CSV

	A	B	C
1	students		
2	student_id	student_name	gpa
3	2538	John Smith	3.5
4	2541	Mary Sue	4
5	2542	Tony Stark	3.8
6			

반정형데이터

2

- HTML
- XML, JSON
- 로그, IoT 센서데이터

```
1 students: [  
2  
3 {  
4   "student_id" : 2538,  
5   "student_name" : "John Smith",  
6   "gpa" : 3.5  
7 },  
8 {  
9   "student_id" : 2541,  
10  "student_name" : "Mary Sue",  
11  "gpa" : 4  
12 },  
13 {  
14   "student_id" : 2542,  
15   "student_name" : "Tony Stark",  
16   "gpa" : 3.8  
17 }  
18 ]  
19
```

비정형데이터

3

- Binary – 동영상, 이미지
- Script : 소셜데이터의 텍스트



Q 만족도



값이 커짐에 따라 만족도가 좋아지고
있다고 할 수 있는 경우 = 순서형 데이터!

EX)

만족도 조사 병의 단계 성적표

서열(순위)척도

구간척도

Q 하루동안 은행에 방문하는 고객의 수



1명, 2명, 3명

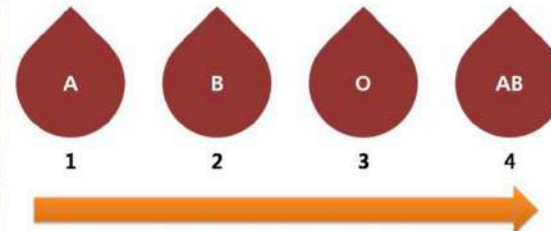
3.5명?
소수점의 의미가
없음

수치적인 의미는 가지고 있으나
소수점으로 표현되지 않는 경우
= 이산형 데이터!

EX)

불량품 수 사고 건수 차량대수

Q 혈액형



값이 달라짐에 따라 좋거나 나쁘다고
할 수 없는 경우 = 명목형 데이터!

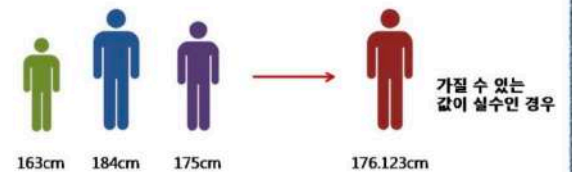
EX)

성별 색깔 취미 주거지역

명목척도

비율척도

Q 키, 몸무게 등



수치적인 의미가 있으며
소수점으로 표현되는 경우
= 연속형 데이터!

EX)

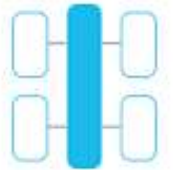
몸무게 시간 길이

SQL vs NoSQL

SQL
Database



Relational

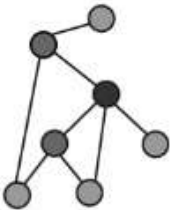


Analytical(OLAP)

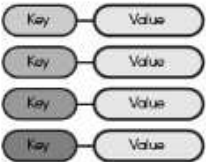
Document



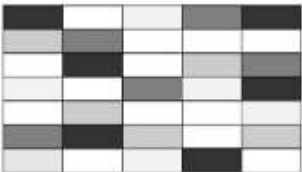
Graph



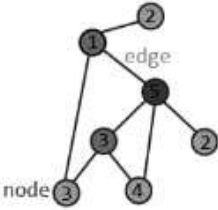
Key-Value



Wide-Column



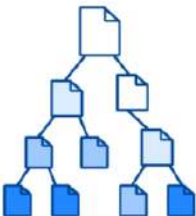
```
{
  "user": {
    "id": "143",
    "name": "improgrammer",
    "city": "New York"
  }
}
```



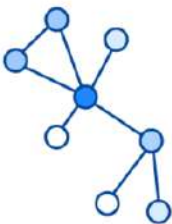
1	Fruit	A Foo	B Baz	
2	City	E DC	D PLA	G FLD
3	State	A NZ	C CL	



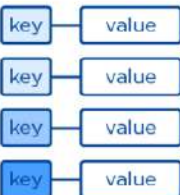
Document



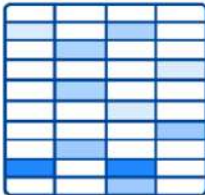
Graph



Key-Value



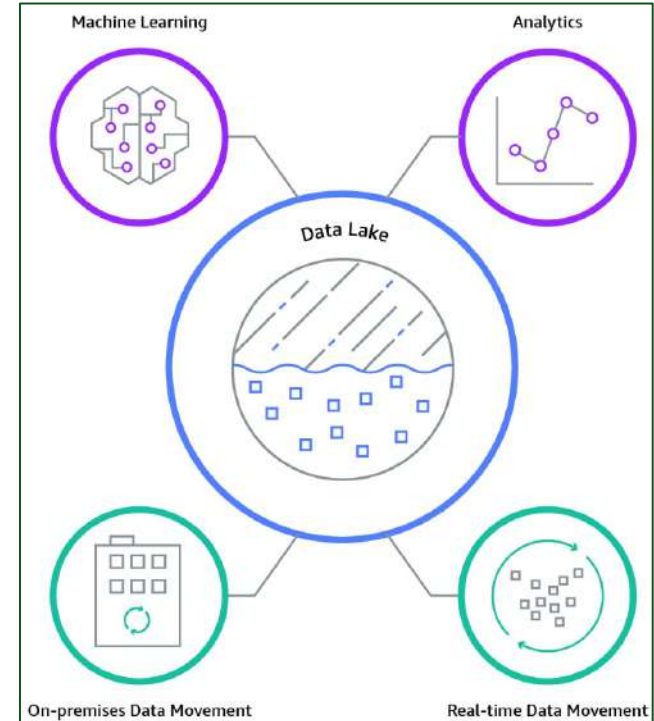
Wide-column



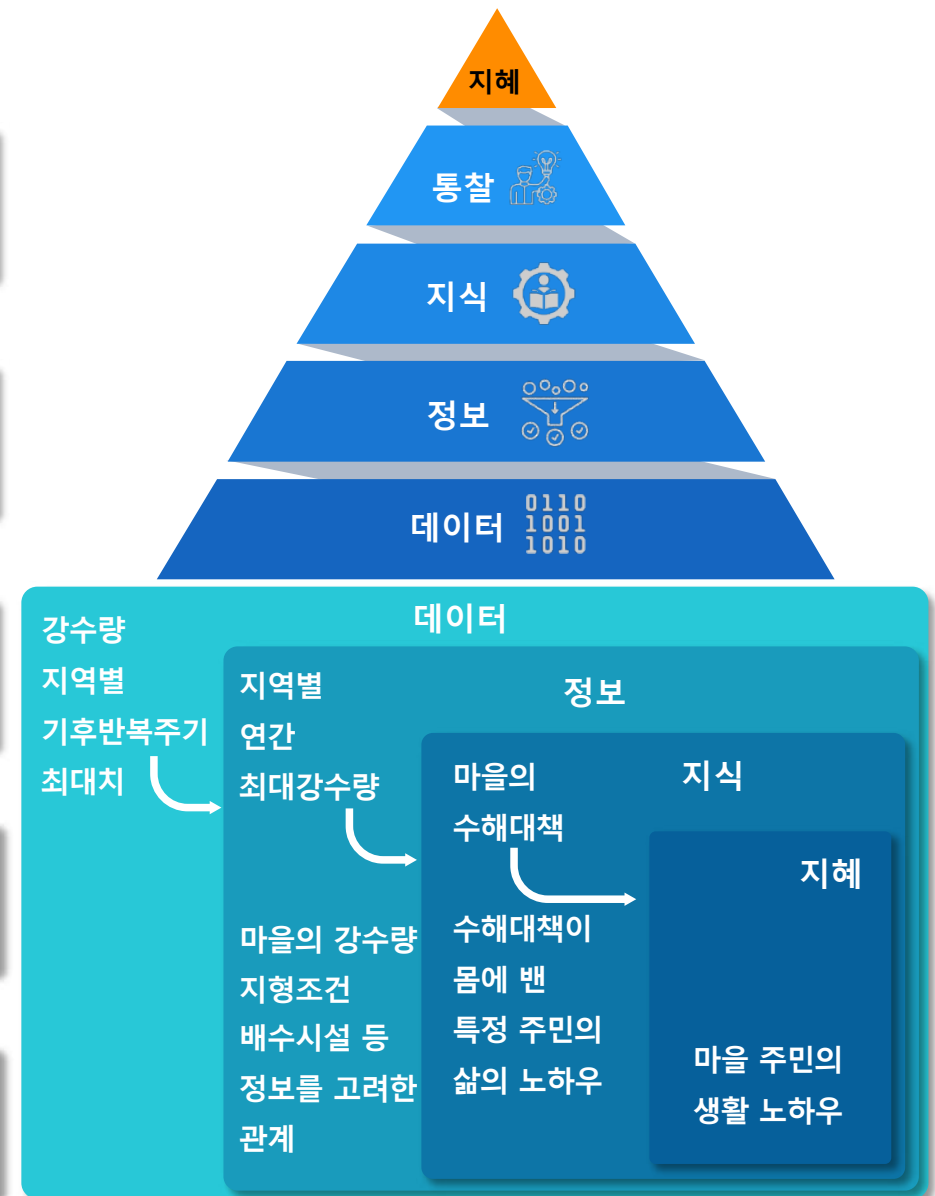
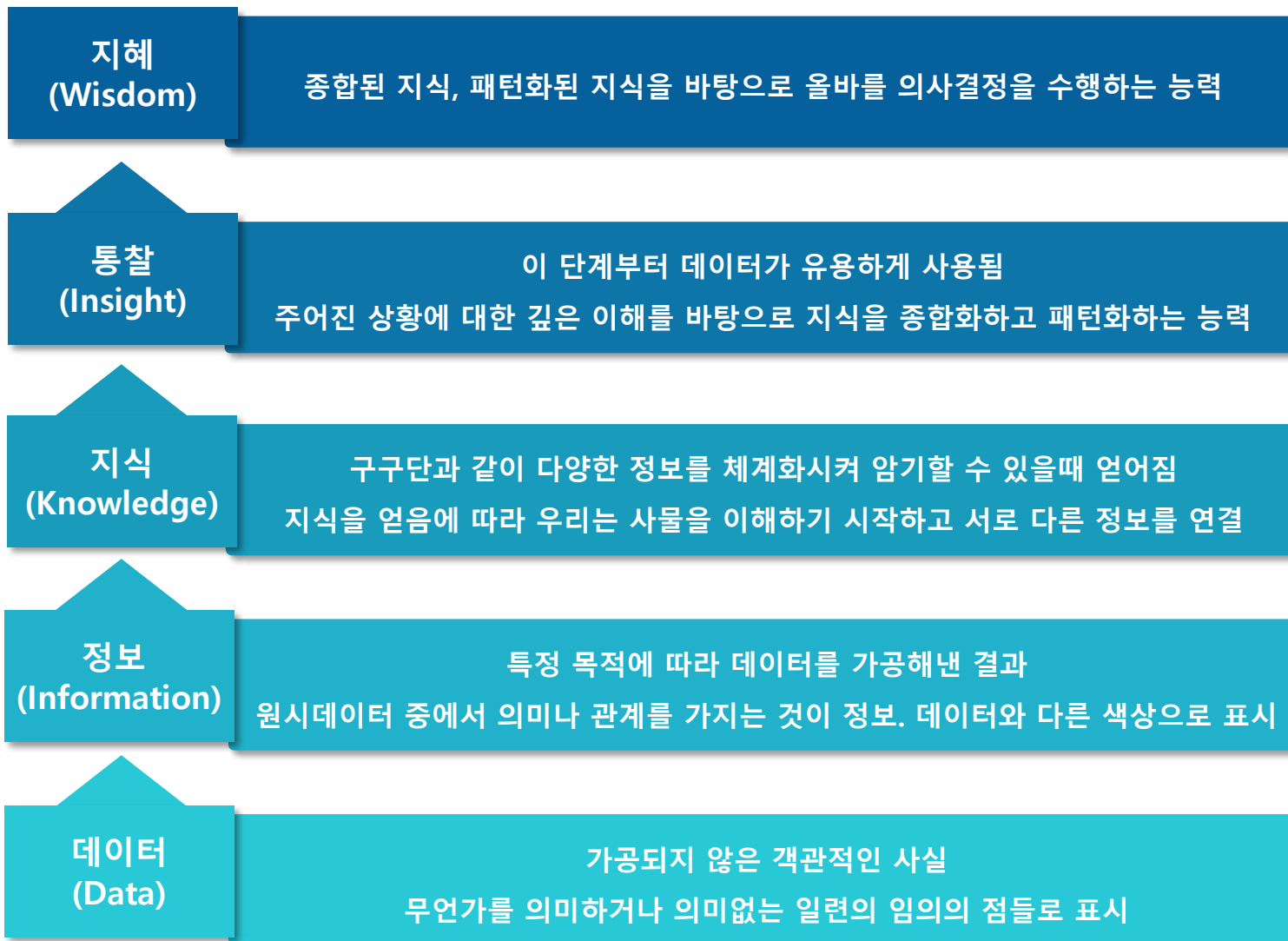
NoSQL
Database

Data Warehouse vs Data Lake

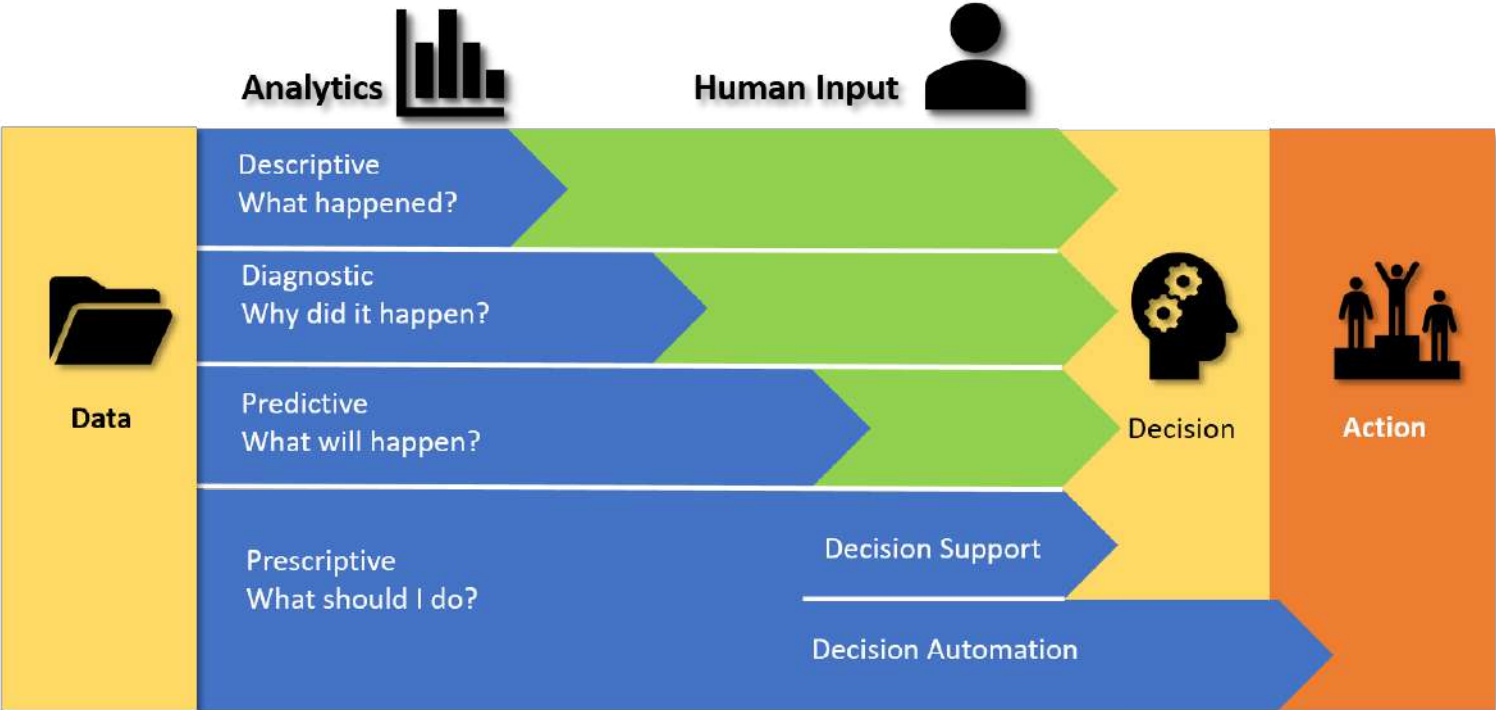
	Data Lake	Data Warehouse
데이터 유형	정형, 반정형, 비정형 데이터	정형데이터
데이터 소스	원시 데이터	정제된 데이터
목적	사전에 정의된 목적 없음	사전에 사용 목적이 정의되어 있음
출처	빅데이터, IoT, 소셜미디어, 스트리밍 데이터	애플리케이션, 비즈니스, 트랜잭션 데이터
사용자	데이터 사이언티스트, 데이터 엔지니어	비즈니스 분석가



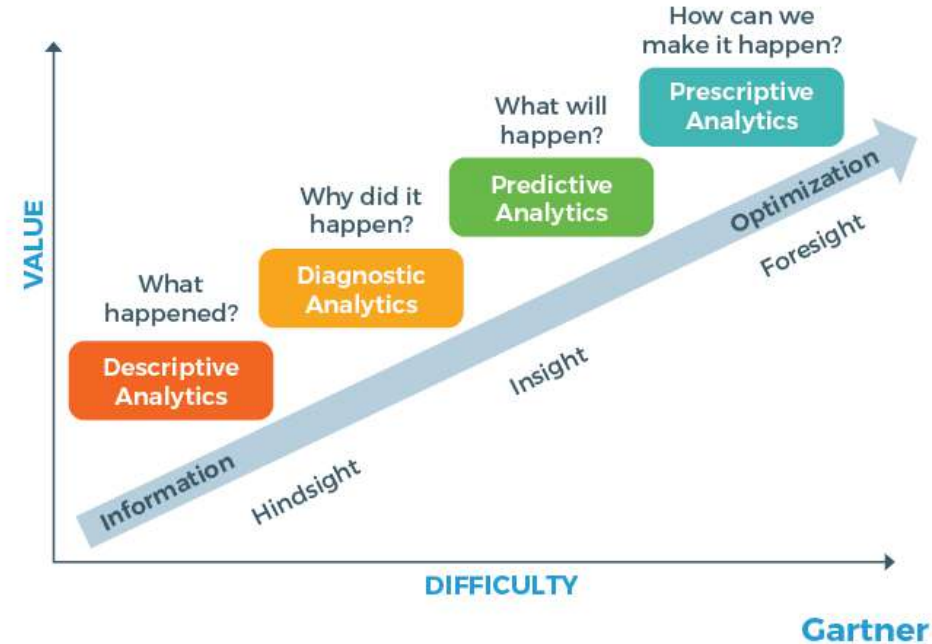
빅데이터 분석 - DIKW



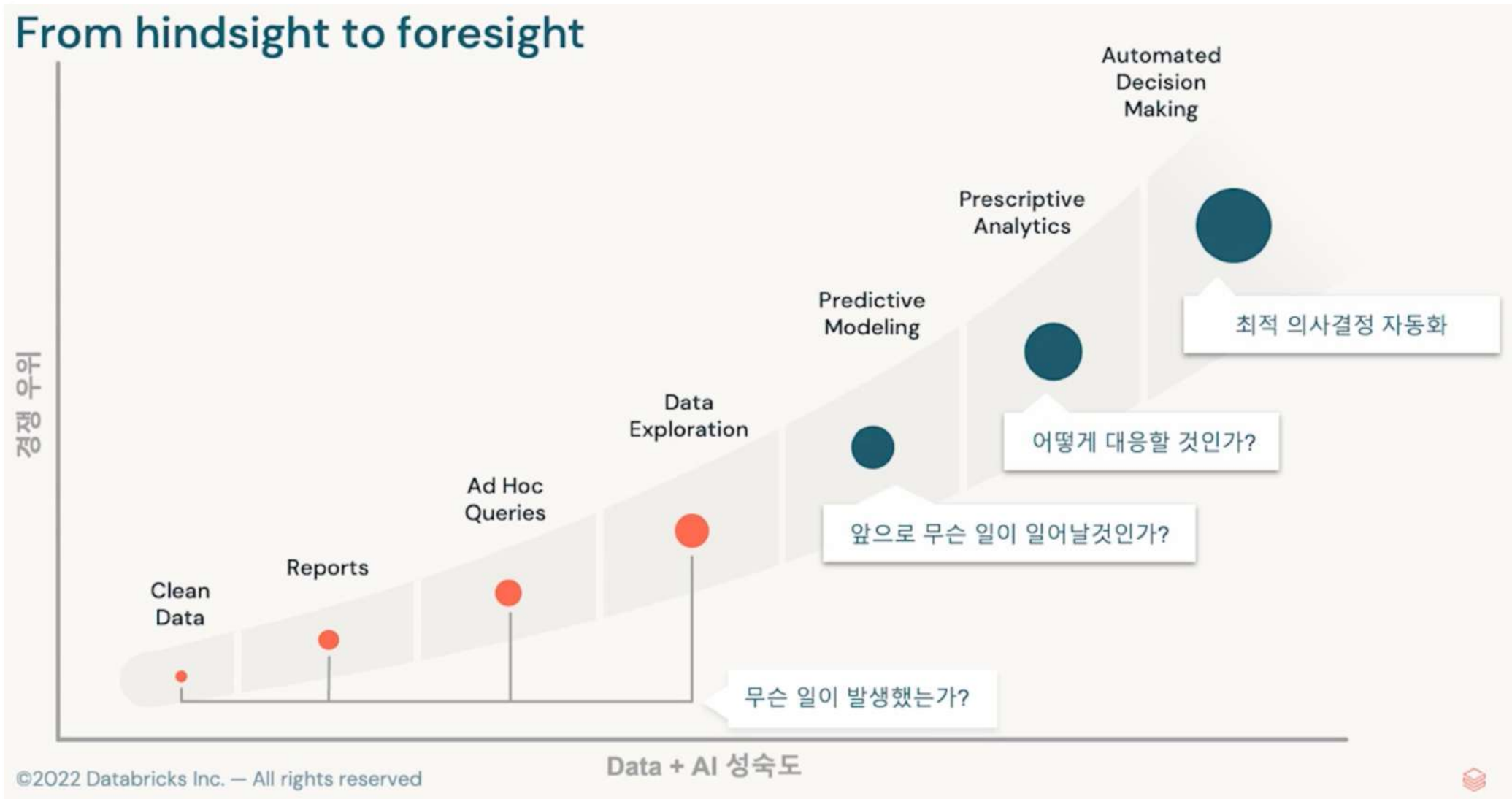
Big Data



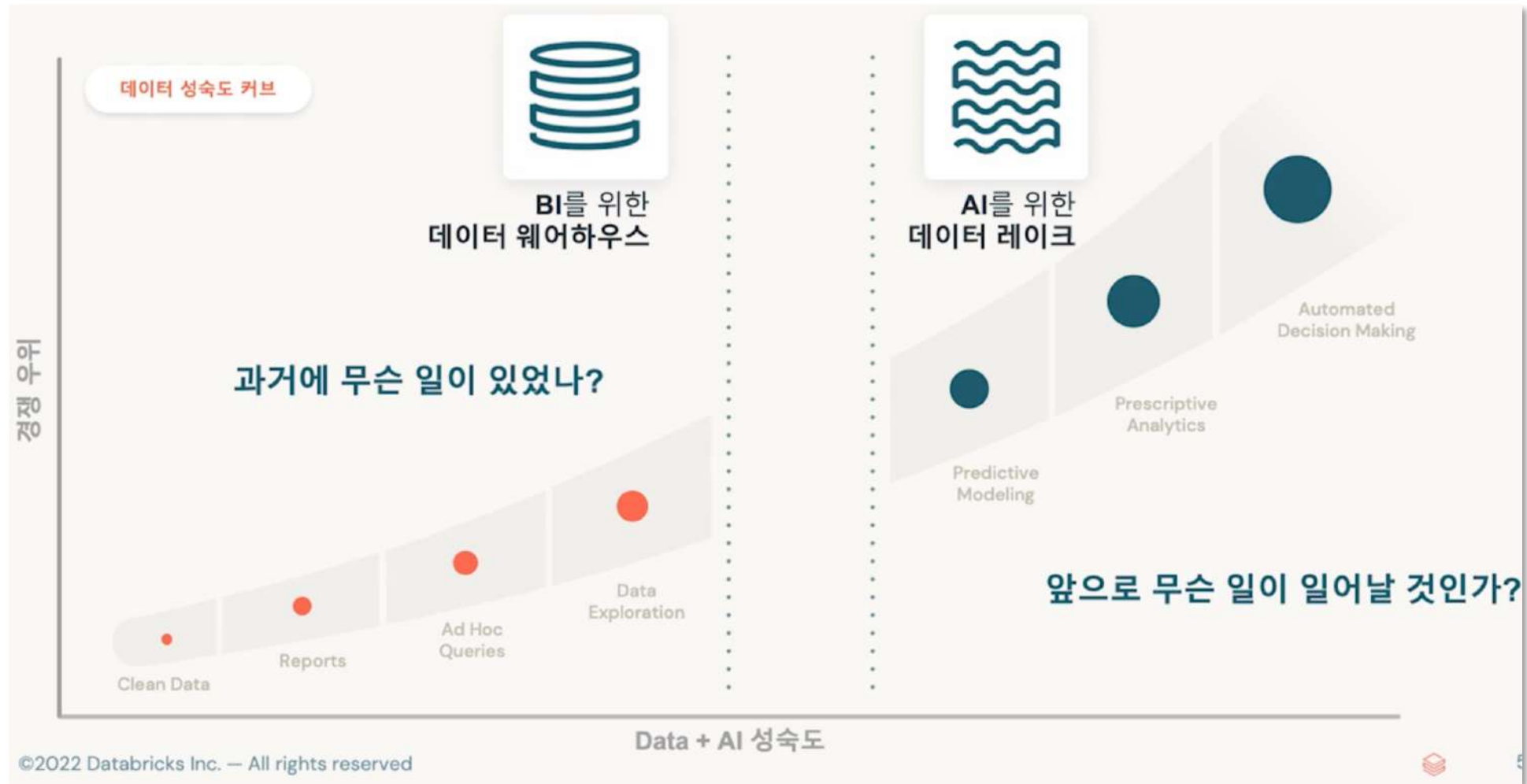
Analytic Value Escalator

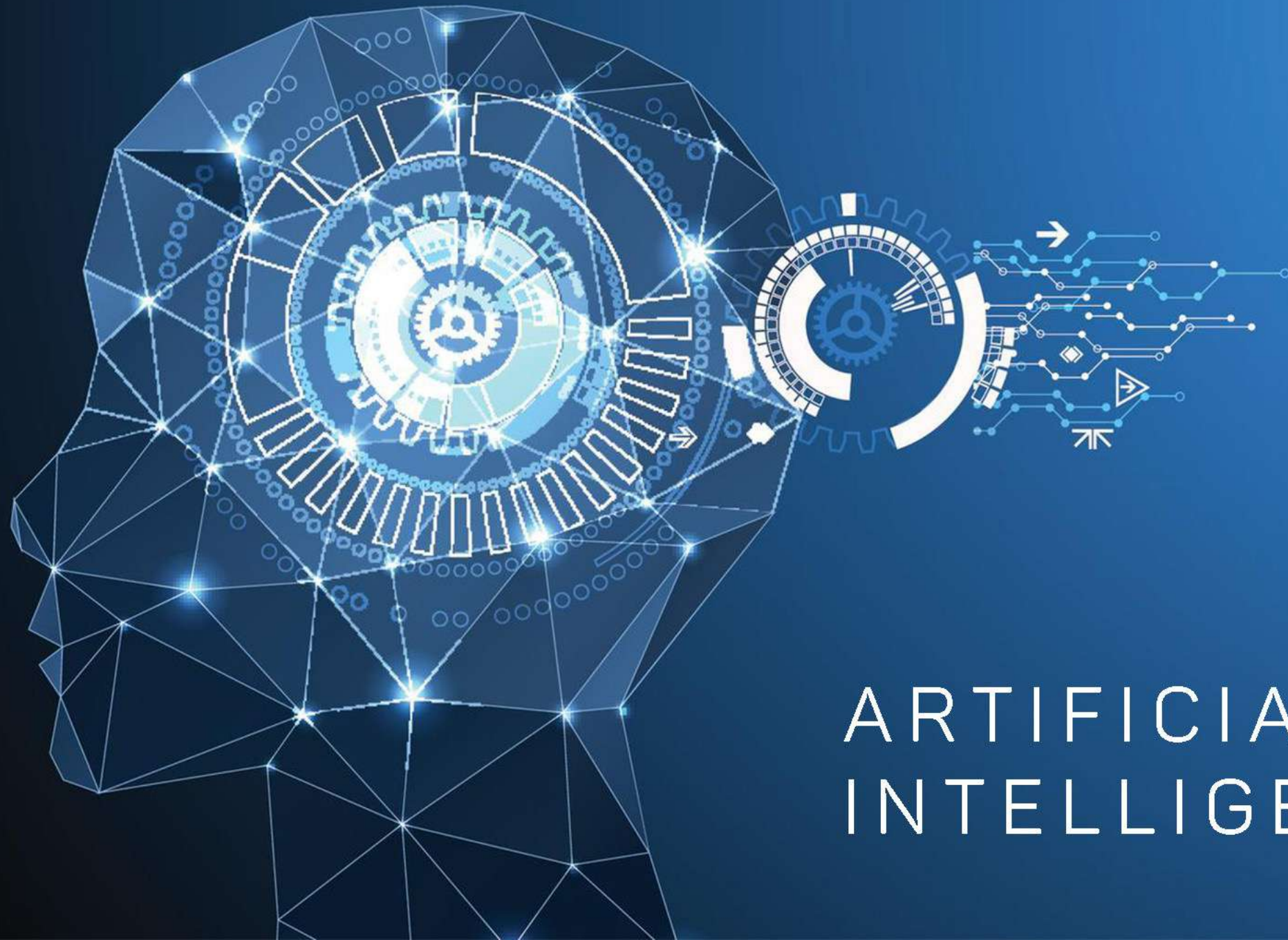


기술 리더들의 데이터 성숙도 커브 위치



목표 달성을 위해서는 서로 다른 두가지 데이터 플랫폼 필요





ARTIFICIAL
INTELLIGENCE



튜링테스트 제2차 세계대전 독일 애니그마 해독

1. 1980년대 전문가 시스템(Expert System)

2. 1997년 IBM 딥블루 vs 개리 카스파로프 체스 게임



3. 2016년 이세돌 vs 알파고



출처: https://live.lge.co.kr/live_with_ai_01/

인공지능의 역사

인공지능(AI)의 역사

1943년

워런 맥클록과 월터 피츠, 전기 스위치처럼 켜고 끄는 기초기능의 인공신경을 그물망 형태로 연결하면 사람의 뇌에서 동작하는 아주 간단한 기능을 흉내낼 수 있음을 증명

1956년



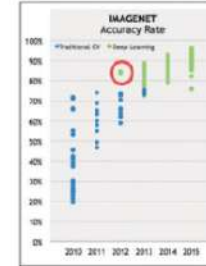
다트머스 회의에서 인공지능 용어 처음 사용. "학습의 모든 면 또는 지능의 다른 모든 특성을 기계로 정확하게 기술할 수 있고 이를 시뮬레이션할 수 있다"

1980년대

전문가들의 지식과 경험을 데이터베이스화해 의사결정 과정을 프로그래밍화한 '전문가 시스템' 도입. 그러나 관리의 비효율성과 유지·보수의 어려움으로 한계

2006년

제프리 힌튼 토론토대 교수, 딥러닝 알고리즘 발표



2012년

국제 이미지 인식 경진대회 '이미지넷'에서 딥러닝 활용한 팀이 우승하며 획기적 전환점

2014년

구글, 딥마인드 인수



1950년

앨런 튜링, 기계가 인간과 얼마나 비슷하게 대화할 수 있는지를 기준으로 기계에 지능이 있는지를 판별하는 튜링 테스트 제안

1958년

프랭크 로센블라트, 뇌신경을 모사한 인공신경 뉴런 '퍼셉트론' 제시

1970년대

AI 연구가 기대했던 결과를 보여주지 못하자 대규모 투자가 중단되며 암흑기 도래

1997년

IBM 딥블루, 체스 챔피언 개리 카스파로프와의 체스 대결에서 승리

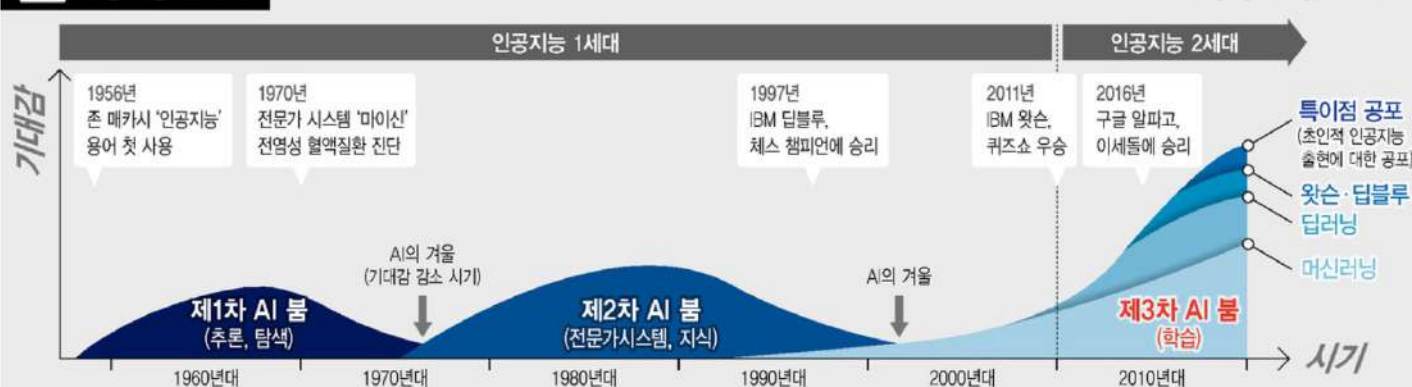
2016년

구글 알파고, 이세돌에게 승리



인공지능의 역사

자료: 마쓰오 유타카의 <인공지능과 딥러닝>, 제프 호킨스



전문가 시스템(Expert System)

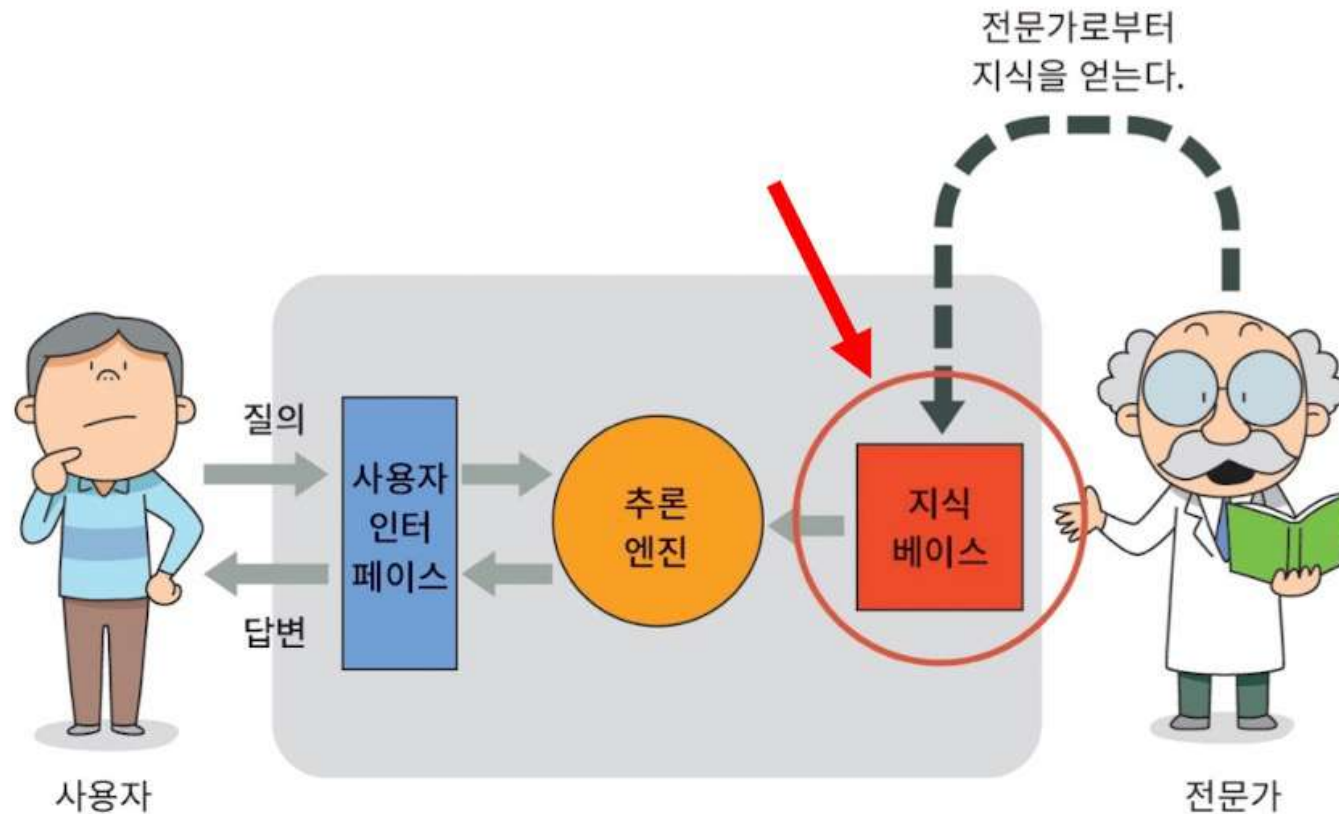
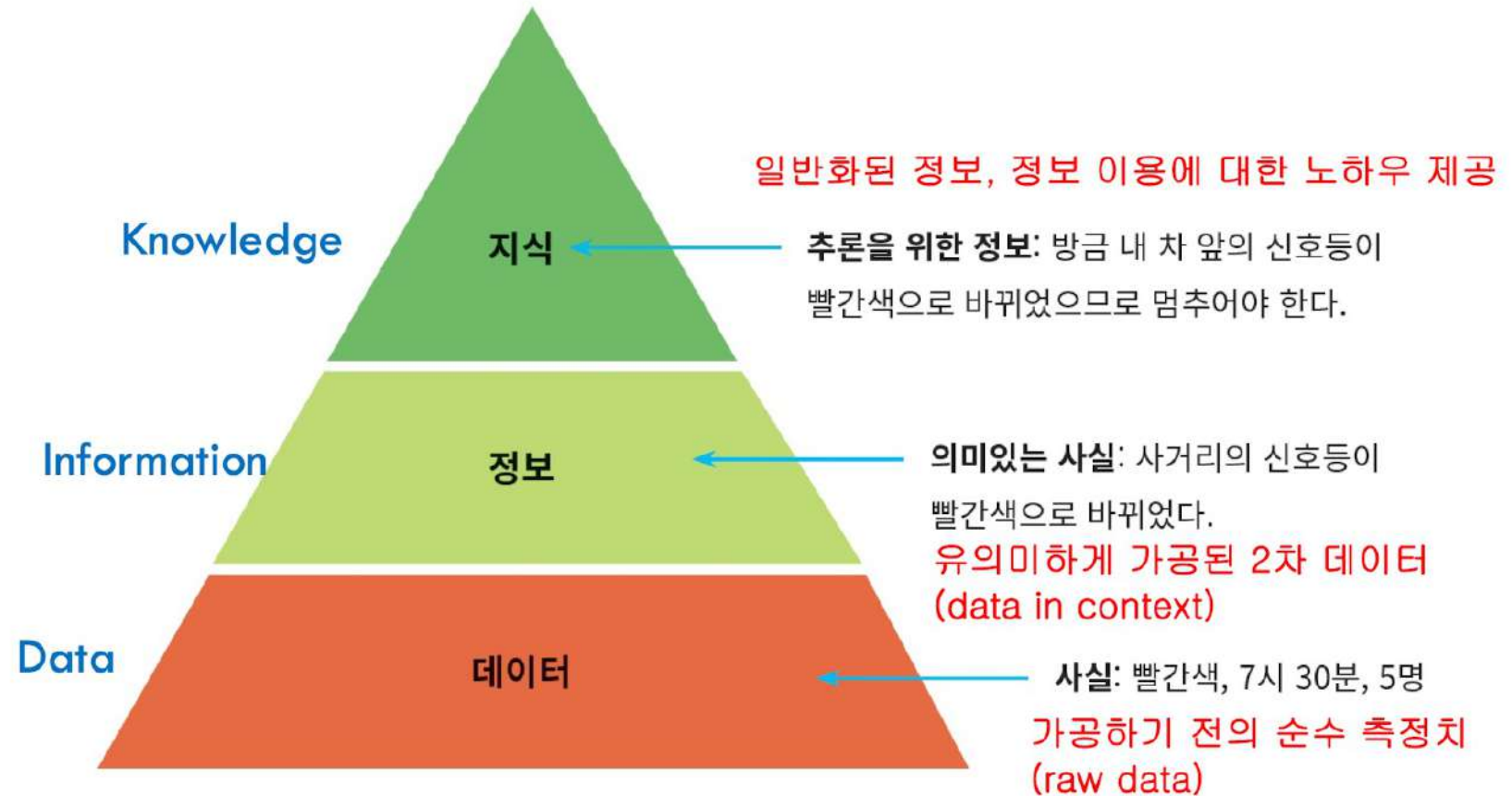
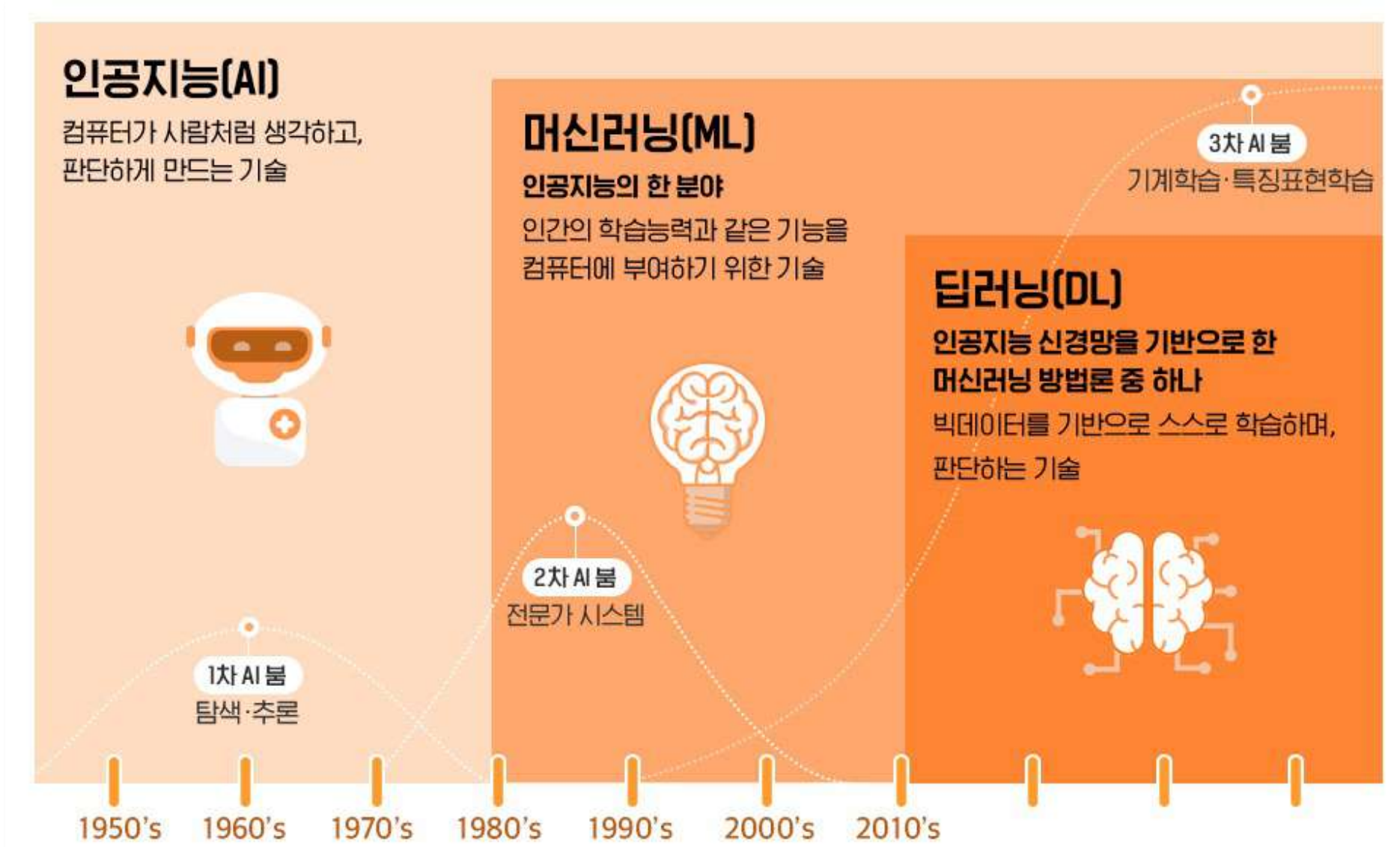


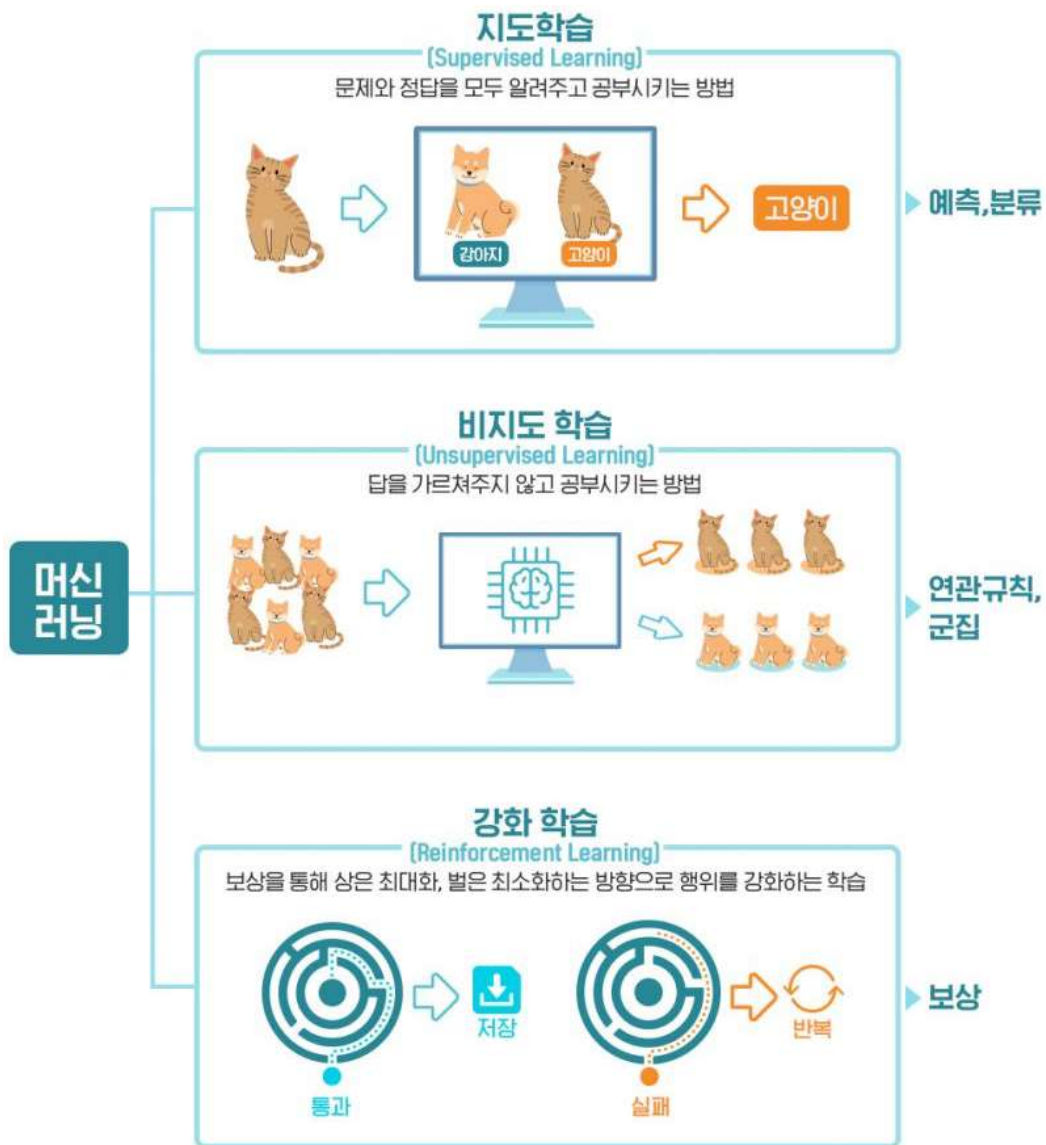
그림 4-4 전문가 시스템과 지식

데이터, 정보, 지식, 지혜





인공지능 분류



① 지도 학습 (Supervised Learning)

: 정답(Label)을 알고 있는 엄마가 아이에게 말을 가르치는 것처럼, 정답을 기반으로 오류를 줄여서 학습하는 방법

② 비지도 학습 (Unsupervised Learning)

: 정답을 모르더라도 유사한 것들과 서로 다른 것들을 구분해서 군집을 만들 수 있는 학습하는 방법.

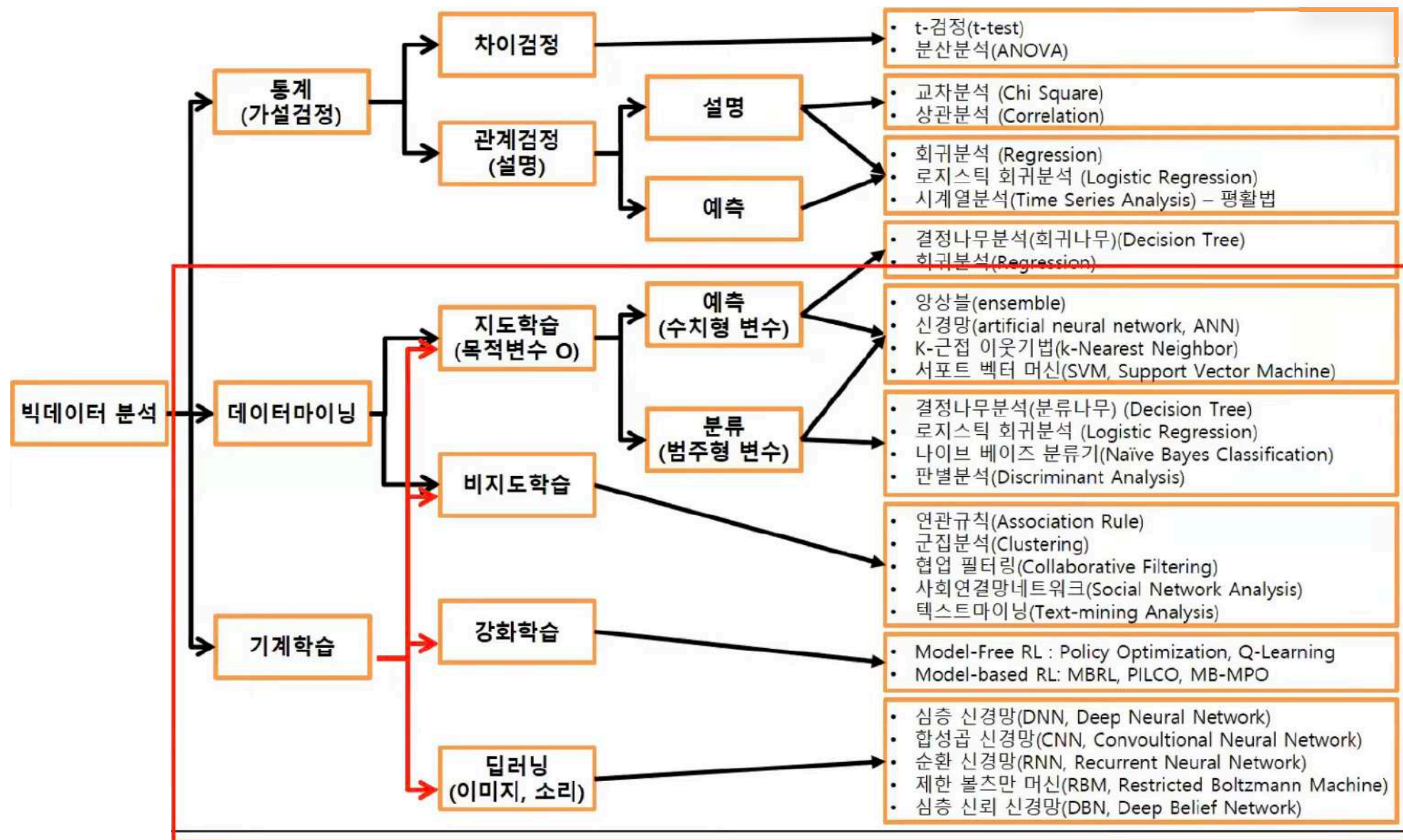
사람이 성장하게 되면 정답을 모르더라도 서로 유사한 데이터들끼리 모아서 군집을 만드는 지능이 생김. 최근에는 지도 학습과 비지도 학습을 섞어서 데이터에 정답을 유추해 나가는 방법의 기술들도 많이 나오고 있으며, 이를 반지도 학습(Semi-Supervised Learning). 군집을 학습한 후에, 군집의 일부 데이터만 사람이 정답을 매겨주면, 그 군집 전체를 사람이 매긴 정답으로 볼 수 있다는 원리

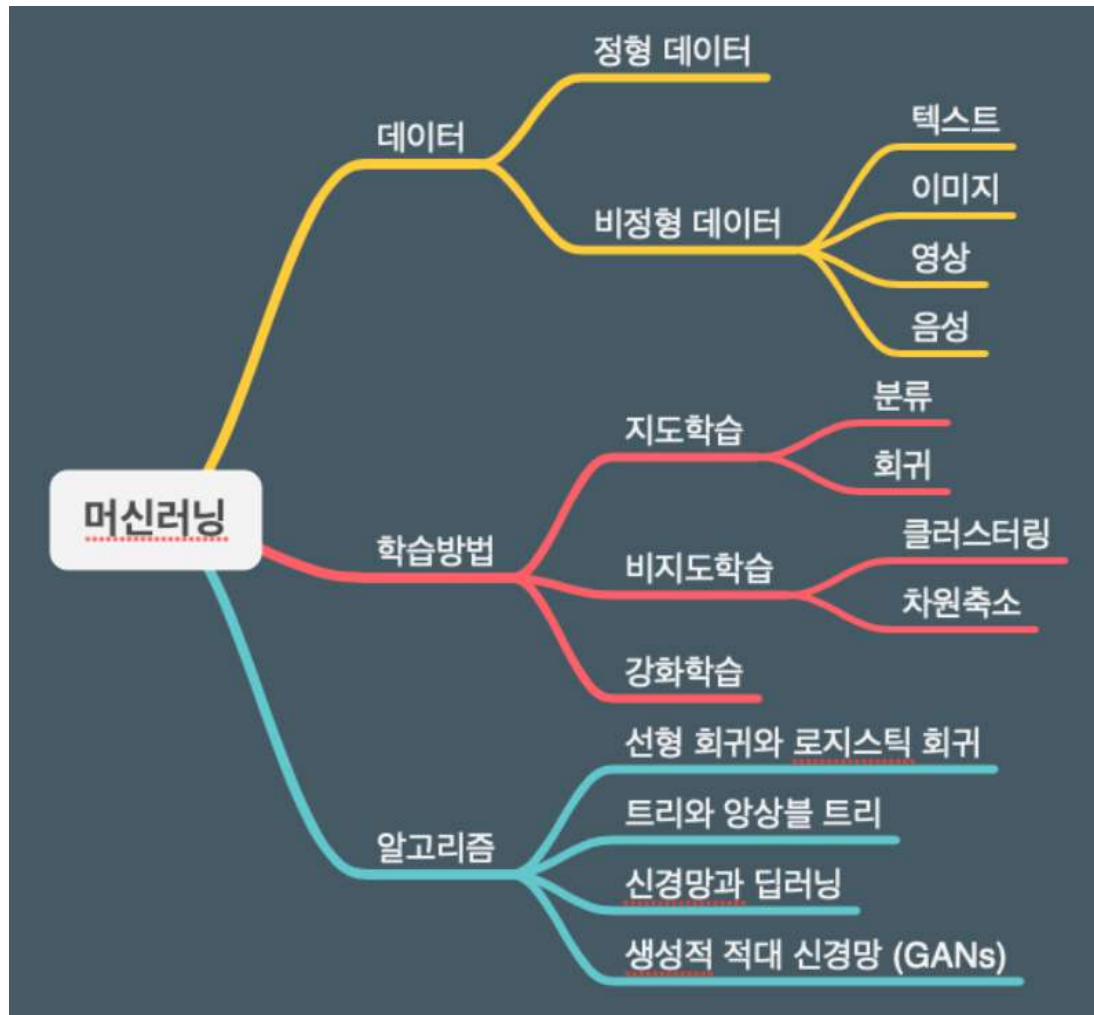
③ 강화 학습 (Reinforcement Learning)

: 보상과 벌칙과 함께 여러 번의 시행착오를 거쳐 스스로 학습하는 방법.

이러한 시행착오에는 2가지 방법이 있음. 이제까지 알아낸 방법 중에 가장 적합한 방법을 시도하는 것과 새로운 방법을 시도해 보는 방법. 기존의 지도학습, 비지도 학습과 달리 보상과 벌칙이 반영된 다음 데이터가 필요하기 때문에 반복 시행을 통해서 데이터를 스스로 만들어가면서 학습해야 함. 일반적으로 많은 학습 시간이 필요해 컴퓨터 시뮬레이션 기반으로 빠르게 반복함으로써 이러한 학습시간의 한계점을 극복

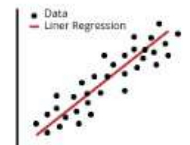
Machine Learning 분석방법



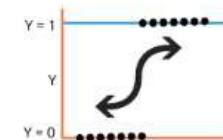


Top 8 Machine Learning Algorithms explained in less than 1 minute each

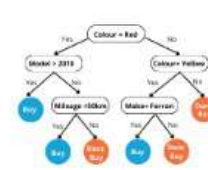
Linear Regression



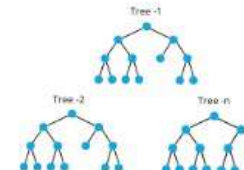
Logistic Regression



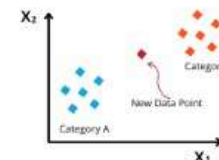
Decision Trees



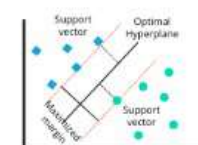
Random Forest



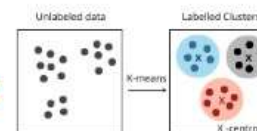
K-Nearest Neighbor



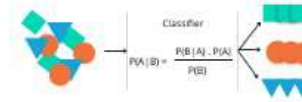
Support Vector Machine



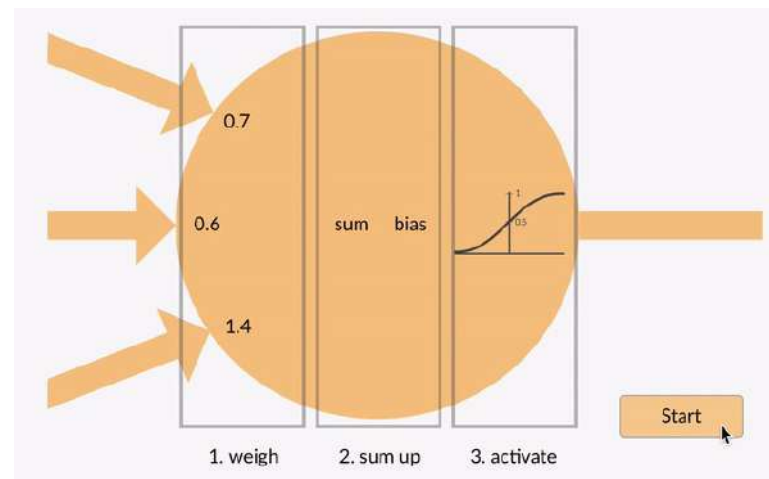
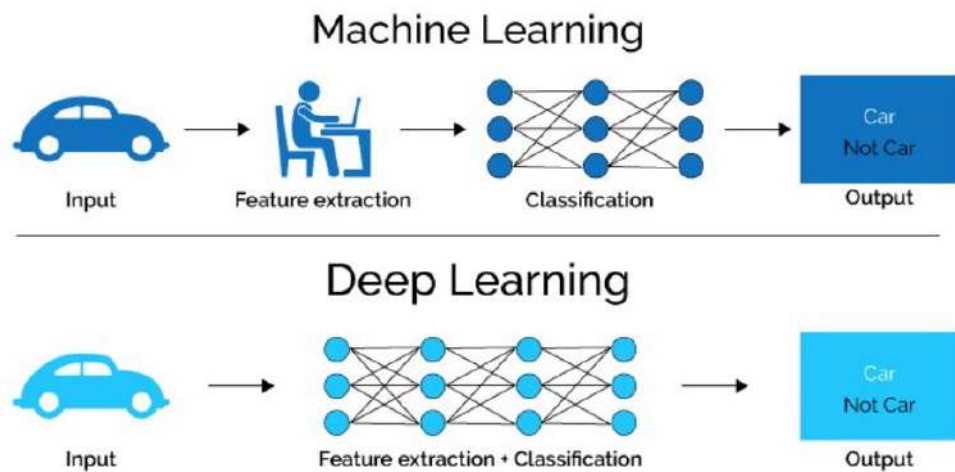
K-Means Clustering



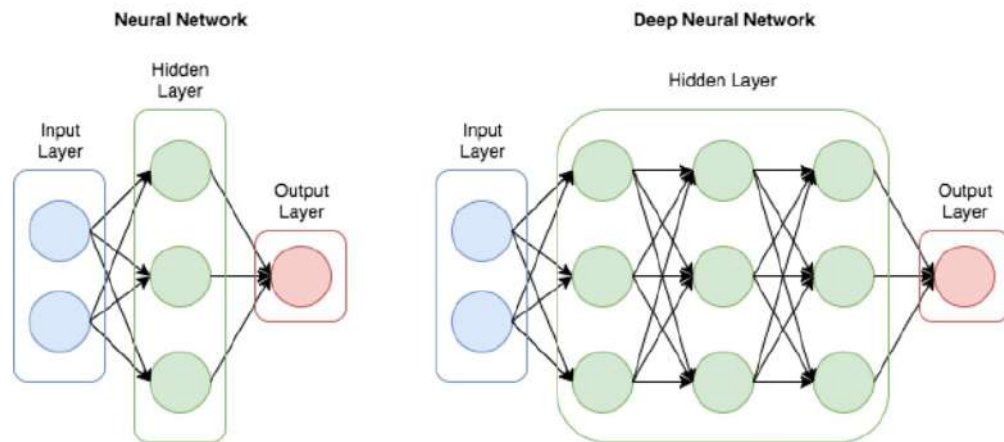
Naïve Bayes



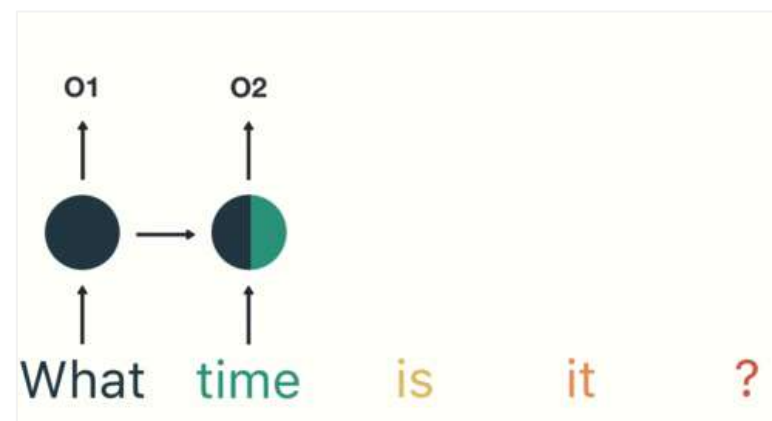
딥러닝 종류



ANN(Artificial Neural Network) - 인공신경망

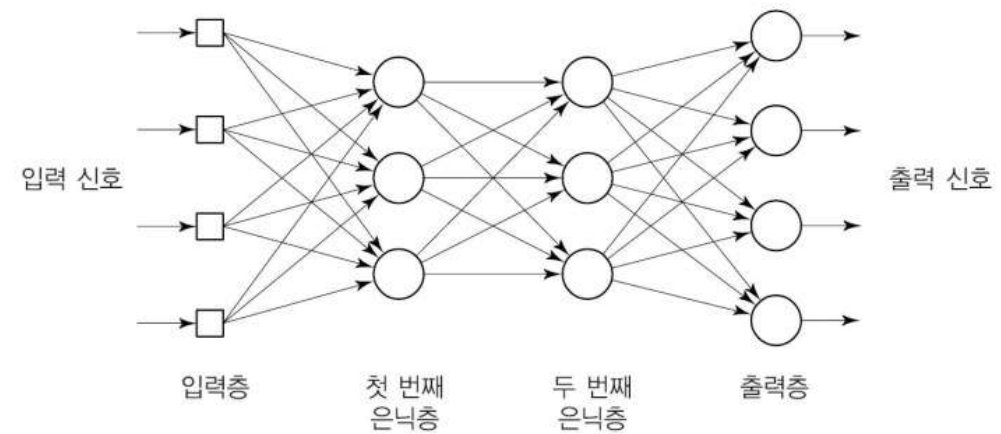
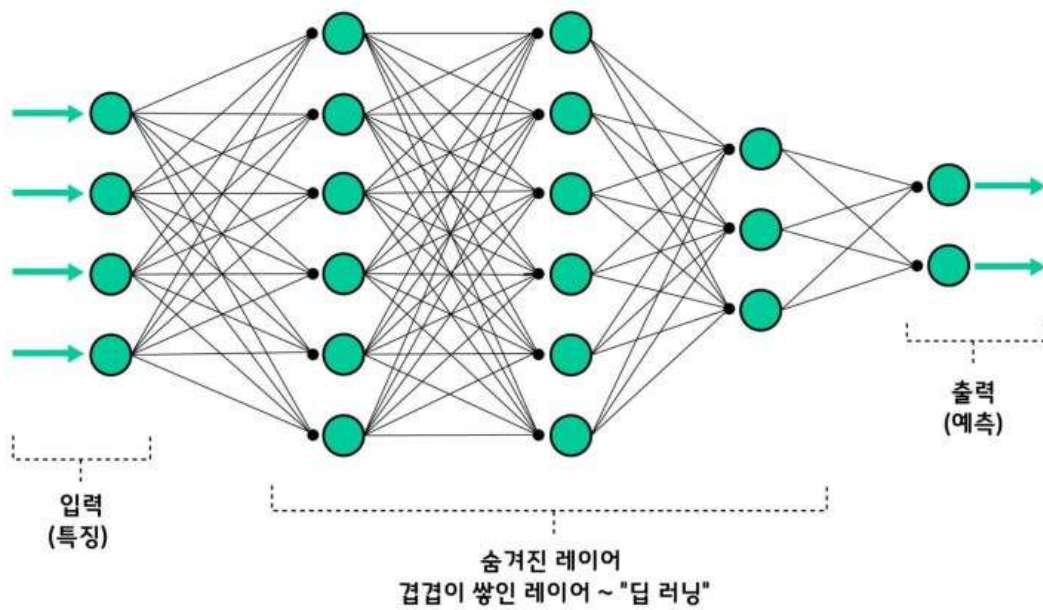


DNN (Deep Neural Network) - 심층신경망



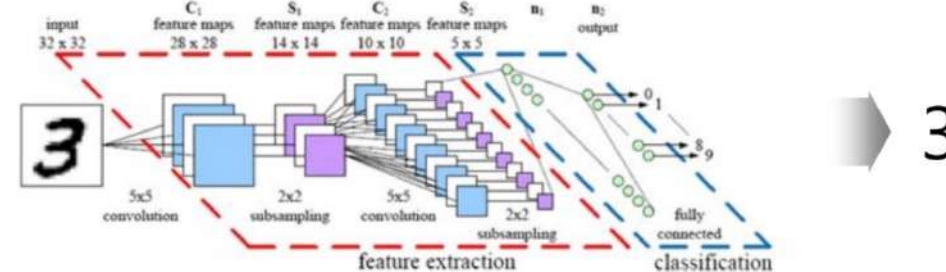
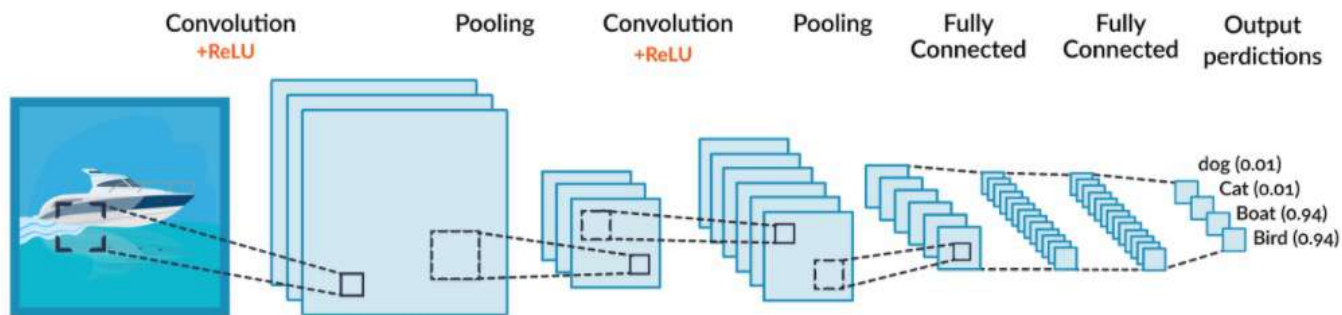
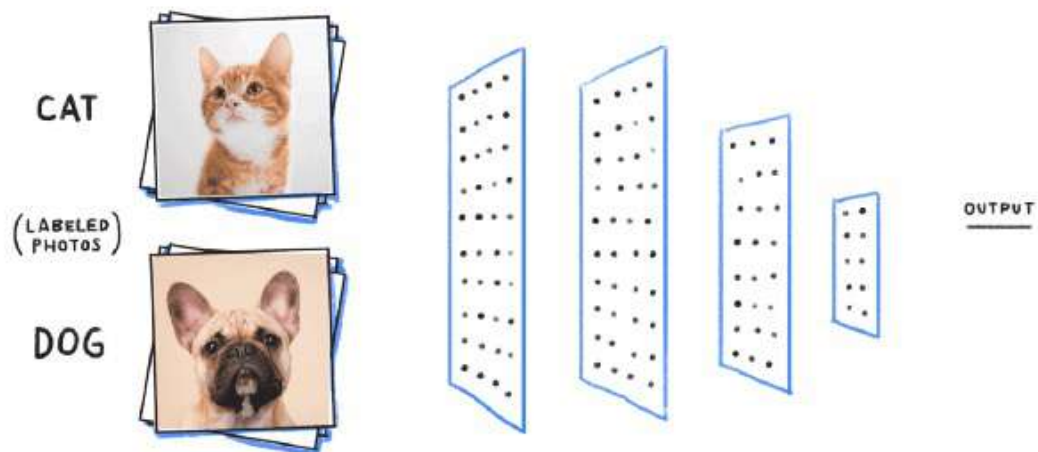
RNN (Recurrent Neural Network) - 순환신경망

딥 러닝 (Deep Learning) - 인공 신경망의 구조



딥러닝 - CNN

CNN(Convolution Neural Network) - 콘볼루션신경망, 합성곱신경망



특징 추출 (Feature Extraction)

- ✓ Convolution layer
 - Pooling (Subsampling)
 - Activation function

분류 (Classification)

- ✓ Fully connected layer
 - Softmax
 - Dropout