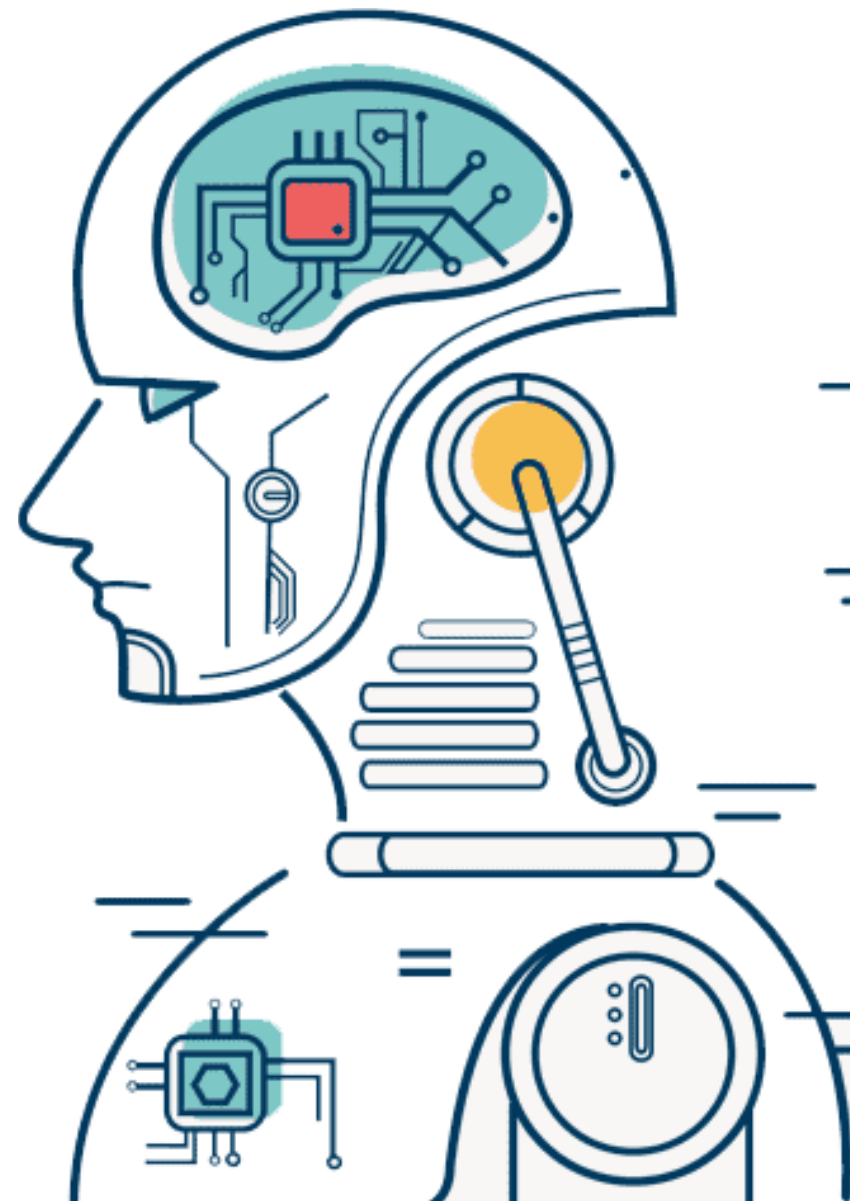


# Machine Learning

## Chapter01 머신러닝 개요

강사 손지영



- Machine Learning 개념을 이해 할 수 있다.
- Machine Learning의 종류 및 과정을 알 수 있다.
- 기계학습과 관련된 기본 용어를 알 수 있다.



Alan Turing  
(1912-1954)  
영국



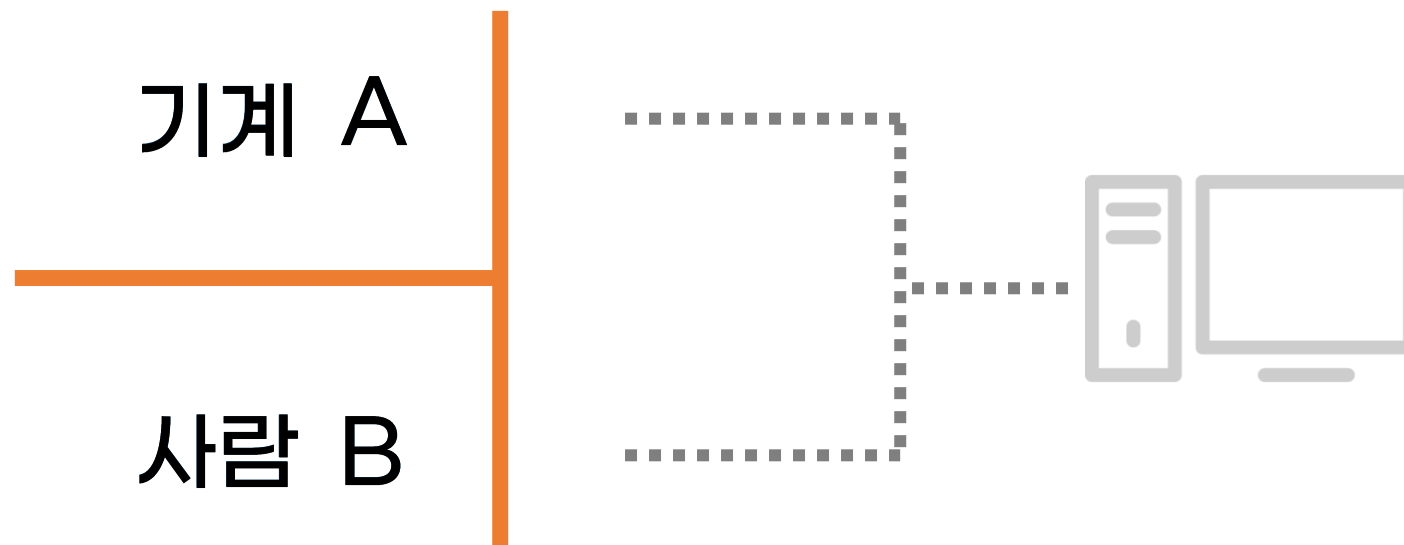
Alan Turing  
(1912-1954)  
영국  
물리학자



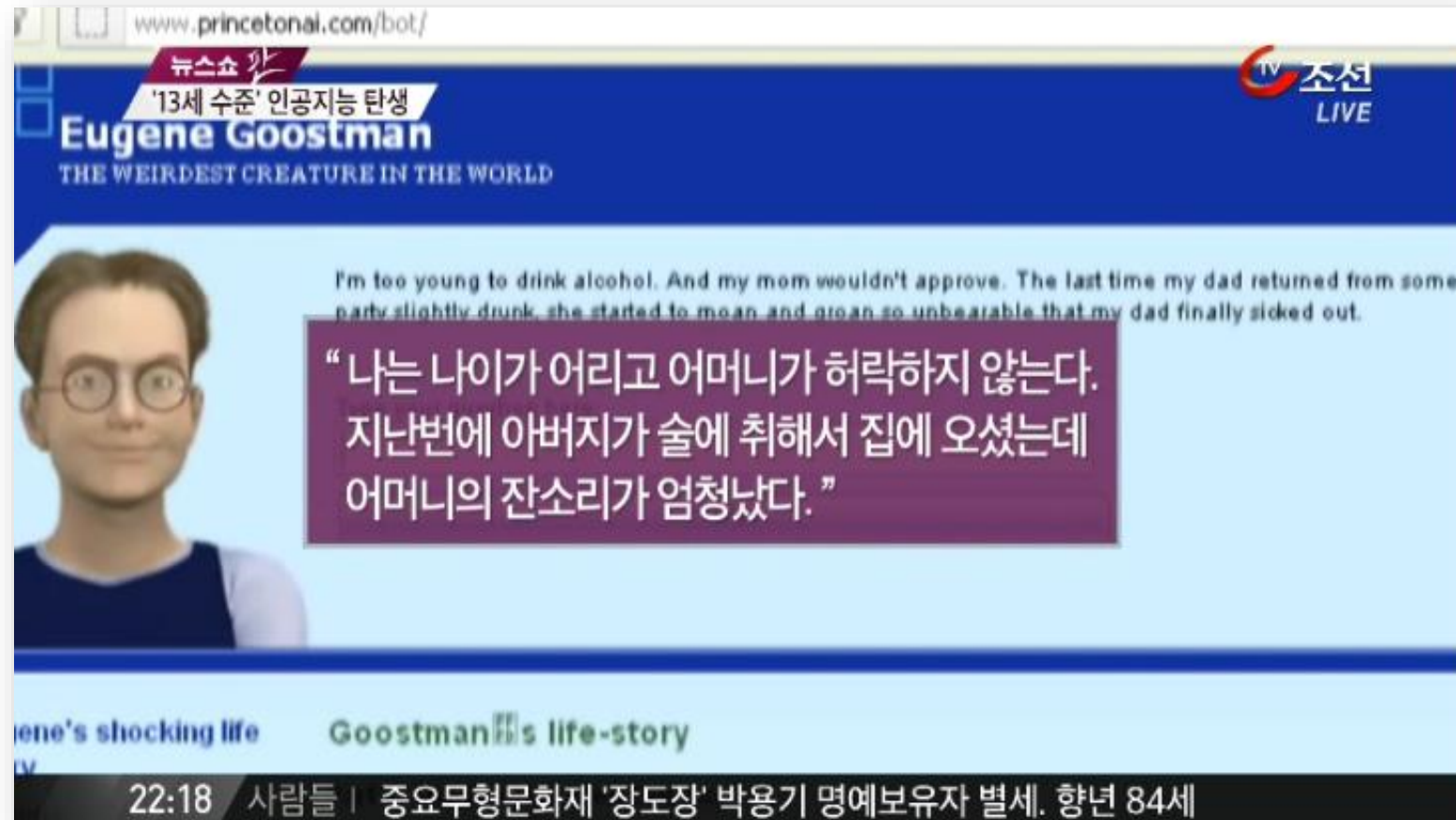
앨런 튜링  
1950년 발표 논문  
'기계도 생각할 수 있을까  
(Can Machines Think?)'

## Turing Test (튜링 테스트)

기계가 인간과 비슷하게 대화가 가능하다면  
기계에 지능이 있다고 인정하는 방식

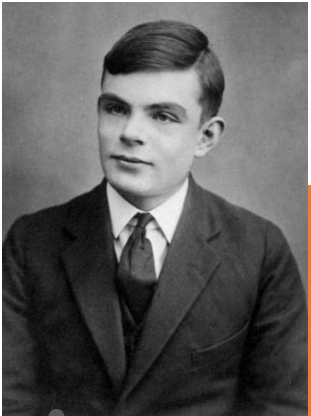


## “Turing Test 최초 통과 유진 구스트만”



# Machine Learning 역사

## Dartmouth AI Project (1956) 인공지능 용어 등장



Alan Turing (1912~1954)  
Turing Test('1950)  
기계(컴퓨터)가 인공지능을  
갖추었는지를 판별하는 실험



암흑기  
1st AI Winter

암흑기  
2nd AI Winter

1950

1960

1970

1980

1990

2000

2010 ~

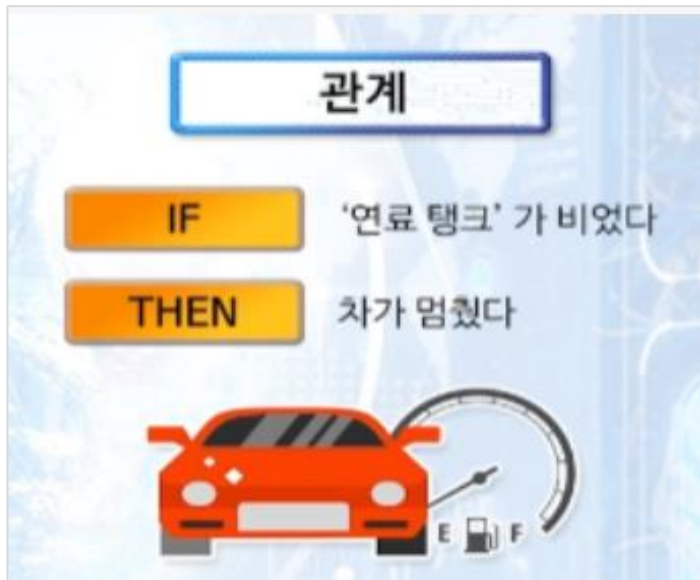
인간 전문가들의 지식을 하나하나  
입력해 만든 규칙 기반 시스템

AI 부활  
Deep Learning



## 규칙 기반 전문가 시스템

“if”와 “else”로 하드 코딩된 명령을 사용하는 시스템





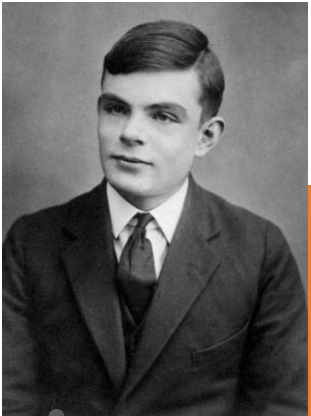
# Rule-based expert system 문제점



- 많은 상황에 대한 규칙들을 모두 만들어 낼 수 없음
- 제작한 로직이 특정 작업에만 국한되어 작업이 조금만 변경되어도 전체 시스템을 다시 만들어야 할 가능성 높음
- 규칙을 설계하려면 해당 분야에 대해서 잘 알아야 함
- 예) 자연어 처리, 스팸 메일 필터, 얼굴 인식 시스템

# Machine Learning 역사

## Dartmouth AI Project (1956) 인공지능 용어 등장



Alan Turing (1912~1954)  
Turing Test('1950)  
기계(컴퓨터)가 인공지능을  
갖추었는지를 판별하는 실험



암흑기  
1st AI Winter

암흑기  
2nd AI Winter

1950

1960

1970

1980

1990

2000

2010 ~

인간 전문가들의 지식을 하나하나  
입력해 만든 규칙 기반 시스템

AI 부활  
Deep Learning

## Beneficial AI 2017 컨퍼런스 [인공지능의 위험성에 대한 우려로 제정]

Ray Kurzweil (Google), Elon Musk (Tesla, SpaceX), Stuart Russell (Berkeley), Demis Hassabis (Deep Mind), Sam Harris (CSER/FLI), Nick Bostrom (FHI), David Chalmers (NYU), Bart Selman (Cornell), Jaan Tallinn (CSER/FLI)



## Beneficial AI 2017 컨퍼런스 [인공지능의 위험성에 대한 우려로 제정]

Question: 초 인공지능이란 영역은 도달 가능한 것인가?

Ray Kurzweil	Elon Musk	Stuart Russell	Demis Hassabis	Sam Harris	Nick Bostrom	David Chalmers	Bart Selman	Jaan Tallinn
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Question: 초 지능을 가진 개체의 출현이 가능할 것이라고 생각하는가?

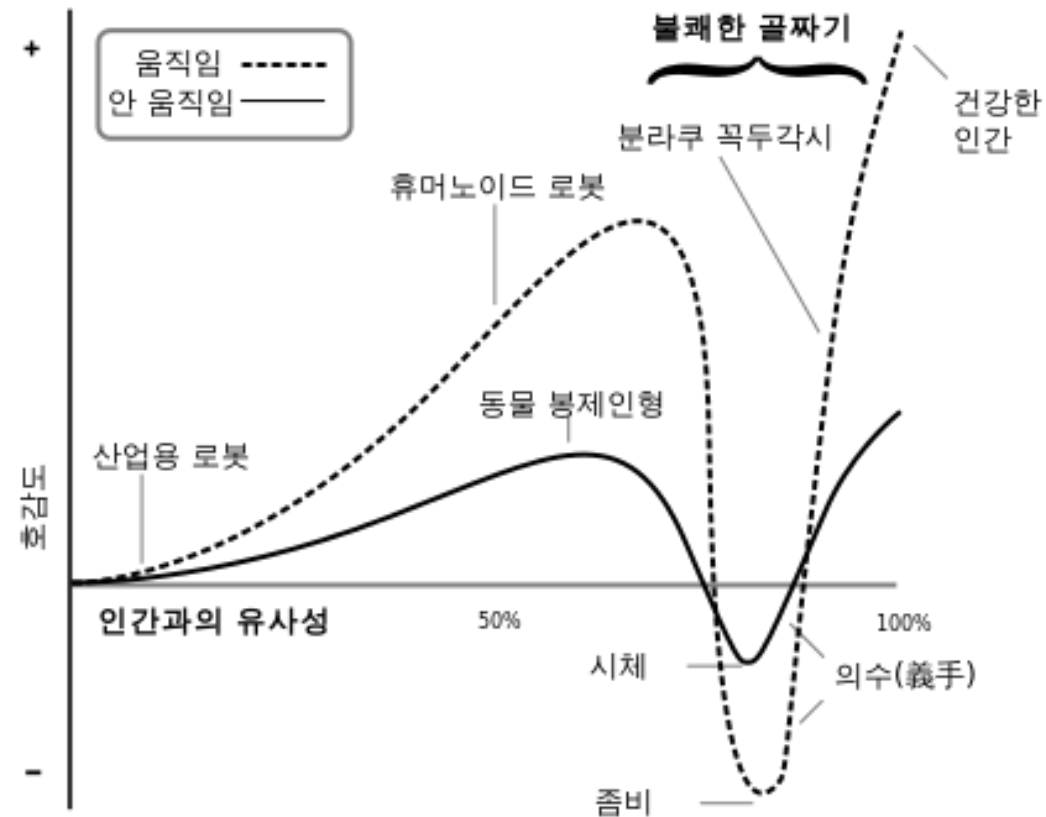
Ray Kurzweil	Elon Musk	Stuart Russell	Demis Hassabis	Sam Harris	Nick Bostrom	David Chalmers	Bart Selman	Jaan Tallinn
Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Question: 초 지능의 실현이 일어나기를 희망하는가?

Ray Kurzweil	Elon Musk	Stuart Russell	Demis Hassabis	Sam Harris	Nick Bostrom	David Chalmers	Bart Selman	Jaan Tallinn
Yes	Complicated	Complicated	Yes	Complicated	Yes	Complicated	Complicated	Complicated

## 불쾌한 골짜기

급격히 강한 거부감을 느끼는 현상



## 불쾌한 골짜기 - 가상 인플루언서



3개월 전

나는 이 광고 너무 트렌디 하고 좋아서 소름돋았는데, 아직은 ai에 사람들이 많이 익숙하지 않아서 거부감 드나범

답글



4개월 전(수정됨)

93년 미드 타임트랙스가 현실로~  
드라마처럼 2199년에 과거로 가는 타임머신이 현실화 되는거 아녀? ㅎㅎ

답글

5개월 전

모델들 되게 신기하게 생겼다 생각했는데 CG였구나

45 답글



5개월 전

나우 대패질하는 캐릭은 불쾌한 골짜기라 불쾌를 넘어 좀 무서움

30 답글

답글 1개



5개월 전

방금 티비 보는데 뭔가 사람이 아닌 거 같아서 바로 찾아봤더니 역시 cg였네.. 와 약간 좀 무섭다.. 저 캐릭터들이 무섭다는게 아니고 우리나라에  
저리로 한다는게... 뭔가 학교에서 배우던.. 미래에는 로봇들이 늘어나서 사람을 일자리가 감소한다 뭐 이렇게 생각나서 무섭네...ㅠㅠ

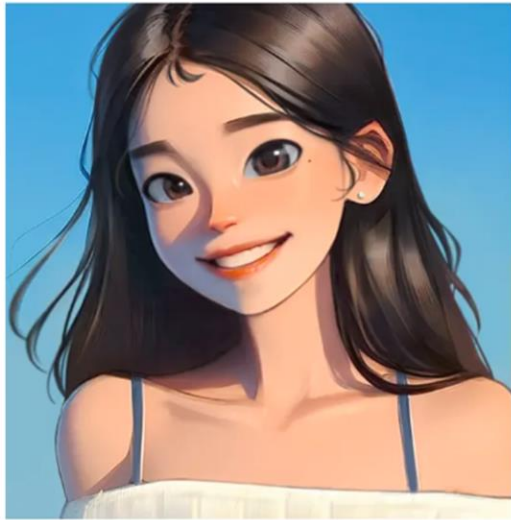
48 답글

답글 1개



## 지능, 윤리&도덕 관점

단순히 똑똑한 것을 넘어 서서 옳고, 그름을 판별할 수 있는 지능을 가지는 것이 인공지능



### ABOUT LUDA

## 내 소개부터 할게!

이름 이루다  
나이 22  
특징 인공지능  
MBTI ENFP  
취미 날씨 좋은날 산책하기, 친구들이  
랑 밤새 수다떨기, 인스타그램 구  
경하기

헉👀  
맛있겠다! 나도 밥 먹고올래  
루다는 뭐 먹게?  
라면 끓여먹을거야!  
ㅋㅋㅋㅋ그래 맛있게 먹어  
신라면 불럭에 계란 탁!  
ㅎㅎ군침 돌아  
귀여워 얼른 끓이고 먹고와!  
내가 라면 팔레시피 찾았어👀  
궁금하지!?

<https://team.luda.ai/ai-ethics>

## 약한인공지능(Artificial Narrow Intelligence)

- 특정분야를 위해 제작된 인공지능
- 체스, 퀴즈, 자율주행, 상품 추천 등

## 강한인공지능(Artificial General Intelligence)

- 모든 방면에서 인간급의 인공지능
- 사고, 계획, 문제해결, 추상화, 복잡한개념학습

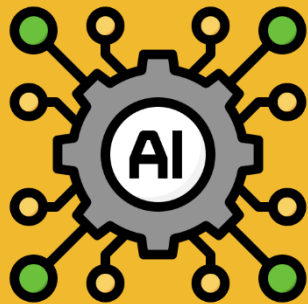
## 초인공지능(Artificial Super Intelligence)

- 과학기술, 사회적 능력 등 모든 영역에서 인간보다 뛰어남
- “충분히 발달한 과학은 마법과 구분할 수 없다” -아서C. 클라크

## Artificial Intelligence

### 인공지능

인간의 학습, 추론, 지각 능력 등을  
인공적으로 컴퓨터 시스템으로  
실현하는 기술의 총칭



## Machine Learning

### 머신러닝

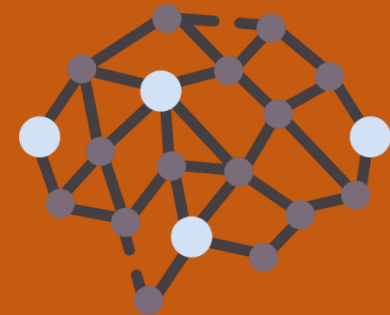
입력된 데이터를 기반으로 기계가  
스스로 학습하여 인공지능 성능을  
향상시키는 기술 방법



## Deep Learning

### 딥러닝

인간의 뉴런과 비슷한 인공신경망  
방식으로 컴퓨터를 스스로 학습하고  
문제를 해결하는 방법





입력 데이터를 이용하여 **특성**과 **패턴**을 찾아  
**학습**하고 **예측**을 수행하는 것

## 머신러닝이 유용한 분야

- 대량의 데이터에서 통찰을 얻어야 하는 문제
- 새로운 데이터에 적응 해야 하는 유동적인 환경
- 기존 솔루션으로는 많은 수동 조정과 규칙이 필요한 문제
- 전통적인 방식으로는 전혀 해결 방법이 없는 복잡한 문제

## 머신러닝으로 할 수 있는 일과 필요한 데이터

머신러닝 수행 기능	데이터(문제)	데이터(실제결과,답)
생존 예측		
집 가격 예측		
긍정/부정 감성 예측		
행동 이상 탐지		
꽃 품종 예측		
숫자 이미지 분류		
관심도 예측, 맞춤 상품 추천		



## 머신러닝 종류

- ✔ 지도학습 (Supervised Learning)
- ✔ 비지도학습 (Unsupervised Learning)
- ✔ 강화학습 (Reinforcement Learning)

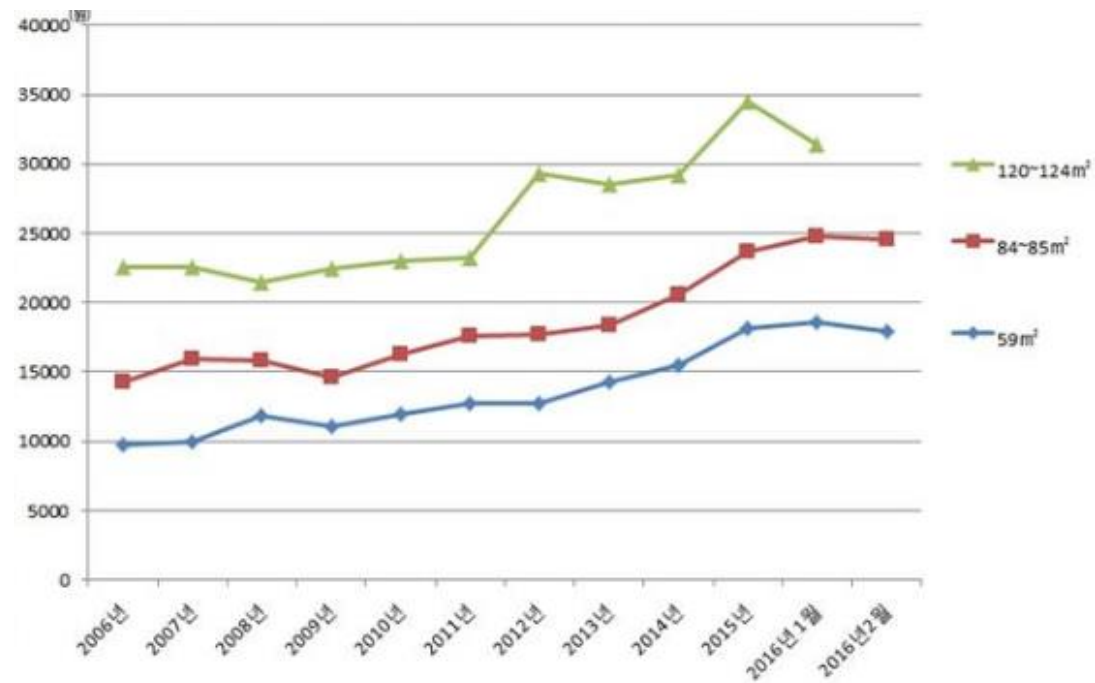
## ✔ 지도학습 (Supervised Learning)

- 데이터에 대한 Label(명시적인 답)이 주어진 상태에서 학습시키는 방법
- 분류(Classification)와 회귀(Regression)

## ✔ 지도학습 (Supervised Learning) 예시



스팸 메일 분류



집 가격 예측

## ✓ 지도학습 (Supervised Learning) - 분류 (Classification)

- 미리 정의된 여러 클래스 레이블 중 하나를 예측하는 것
- 속성 값을 입력, 클래스 값을 출력으로 하는 모델
- 암 양성 음성 중 하나로 분류, 붓꽃(iris)의 세 품종 중 하나로 분류 등
- 이진분류, 다중 분류

## ✓ 지도학습 (Supervised Learning) - 회귀 (Regression)

- 연속적인 숫자를 예측하는 것
- 속성 값을 입력, 연속적인 실수 값을 출력으로 하는 모델
- 어떤 사람의 교육수준, 나이, 주거지를 바탕으로 연간 소득 예측 등
- 예측 값의 미묘한 차이가 크게 중요하지 않음

## ✓ 비지도학습 (Unsupervised Learning)

- 데이터에 대한 Label(명시적인 답)이 없는 상태에서 학습시키는 방법
- 데이터의 숨겨진 특징, 구조, 패턴 등 데이터의 성격을 파악하는데 사용
- 클러스터링(Clustering), 차원축소(Dimensionality Reduction) 등

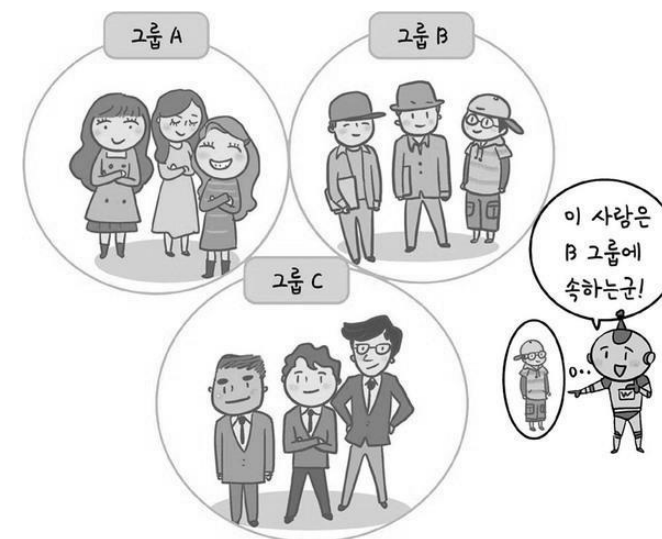


## 비지도학습 (Unsupervised Learning)



한즈온 머신러닝 도서 중

이미지 색상 분할



Copyright © Gilbut, Inc. All rights reserved.

<https://thebook.io/080244/part01/unit02/02-03/>

소비자 그룹 발견을 통한 마케팅

## ✓ 군집화(Clustering) vs. 분류(Classification)

**군집화  
(Clustering)**

**vs.**

**분류  
(Classification)**

대상들을 구분해서 그룹을 만든 것

대상이 어떤 그룹에 속하는 지 판단하는 것

## 강화학습 (Reinforcement Learning)

- 지도학습과 비슷하지만 완전한 답(Label)을 제공하지 않음
- 기계는 더 많은 보상을 얻을 수 있는 방향으로 행동을 학습
- 주로 게임이나 로봇을 학습시키는데 많이 사용

## ✔ 강화학습 (Reinforcement Learning) 예시



## 머신러닝 과정

1. Problem Identification(문제정의)
2. Data Collect(데이터 수집)
3. Data Preprocessing(데이터 전처리)
4. EDA(탐색적 데이터분석)
5. Model 선택, Hyper Parameter 조정
6. Training(학습)
7. Evaluation(평가)

## 머신러닝 과정 1. Problem Identification(문제정의)

- 비즈니스 목적 정의모델을 어떻게 사용해 이익을 얻을까?
- 현재 솔루션의 구성 파악
- 지도 vs 비지도 vs 강화
- 분류 vs 회귀



## Q. 분류? 회귀?

회귀



다음 학기 성적 점수를 예측하려면  
어떤 특성이 필요할까?

직전 학기 성적, 알바 진행 여부, 연애,  
잠자는 시간, 출석률, 학교와 집 사이 거리

## 머신러닝 과정 2. Data Collect(데이터 수집)

- File (CSV, XML, JSON)
- Database
- Web Crawler (뉴스, SNS, 블로그)
- IoT 센서를 통한 수집
- Survey

## 머신러닝 과정 3. Data Preprocessing(데이터 전처리)

- 결측치, 이상치 처리
- Feature Engineering (특성공학): Scaling (단위 변환), Encoding (범주형 -> 수치형), Binning (수치형 -> 범주형), Transform (새로운 속성 추출)

## 머신러닝 과정 4. EDA(탐색적 데이터분석)

- 기술통계, 변수간 상관관계
- 시각화: pandas, matplotlib, seaborn
- Feature Selection (사용할 특성 선택)

## 머신러닝 과정 5. Model 선택, Hyper Parameter 조정

- 목적에 맞는 적절한 모델 선택
- KNN, SVM, Linear Regression, Ridge, Lasso, Decision Tree, Random forest, CNN, RNN ...
- Hyper Parameter  
model의 성능을 개선하기위해 사람이 직접 넣는 parameter

## 머신러닝 과정 6. Model Training(학습)

- `model.fit(X_train,y_train)`  
train 데이터와 test 데이터를 7:3 정도로 나눔
- `model.predict (X_test)`

## 트레인,테스트 데이터

X  
Feature, 독립변수

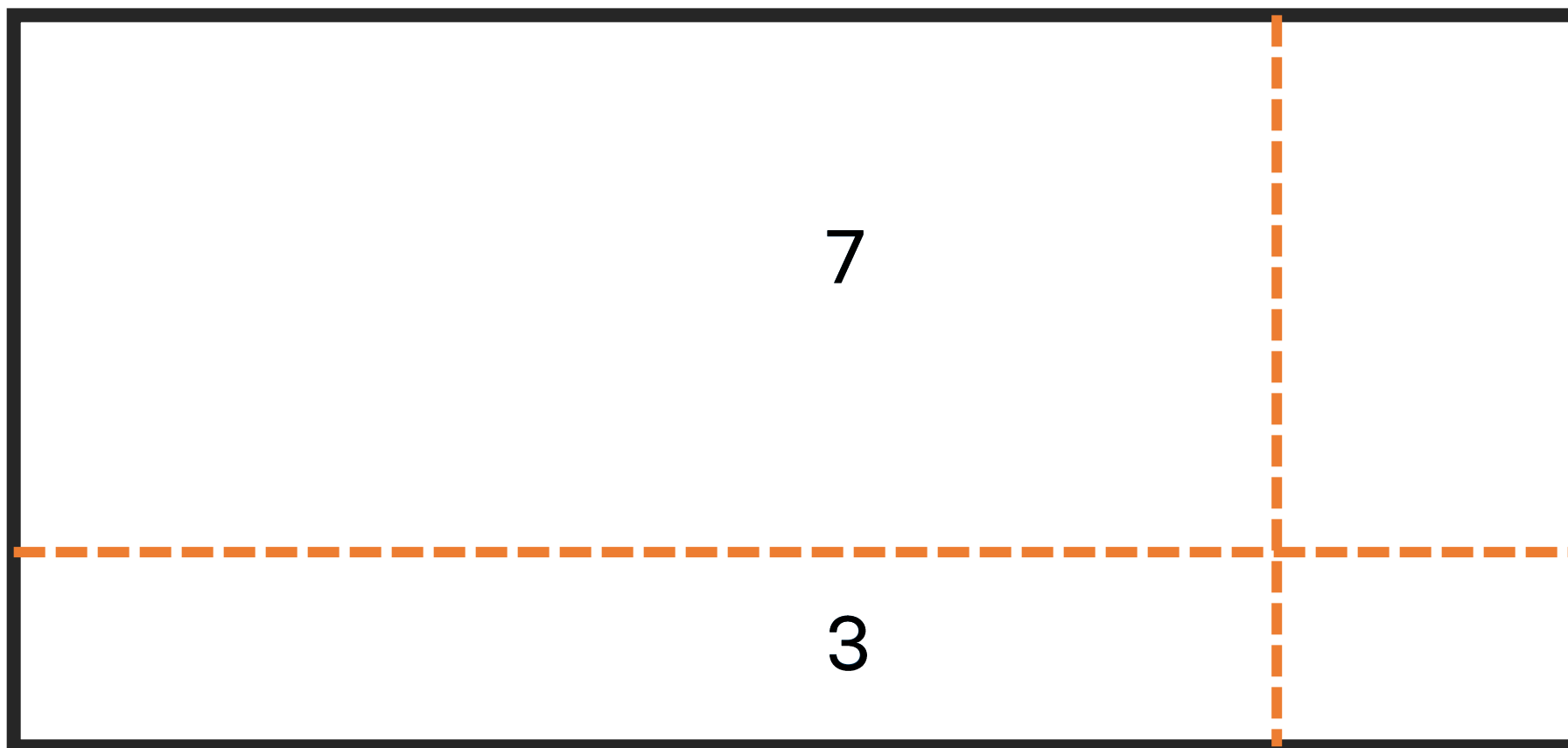
Y  
Label, 종속변수

Train  
Data

7

Test  
Data

3



## 머신러닝 과정 7. Evaluation(평가)

### 분류

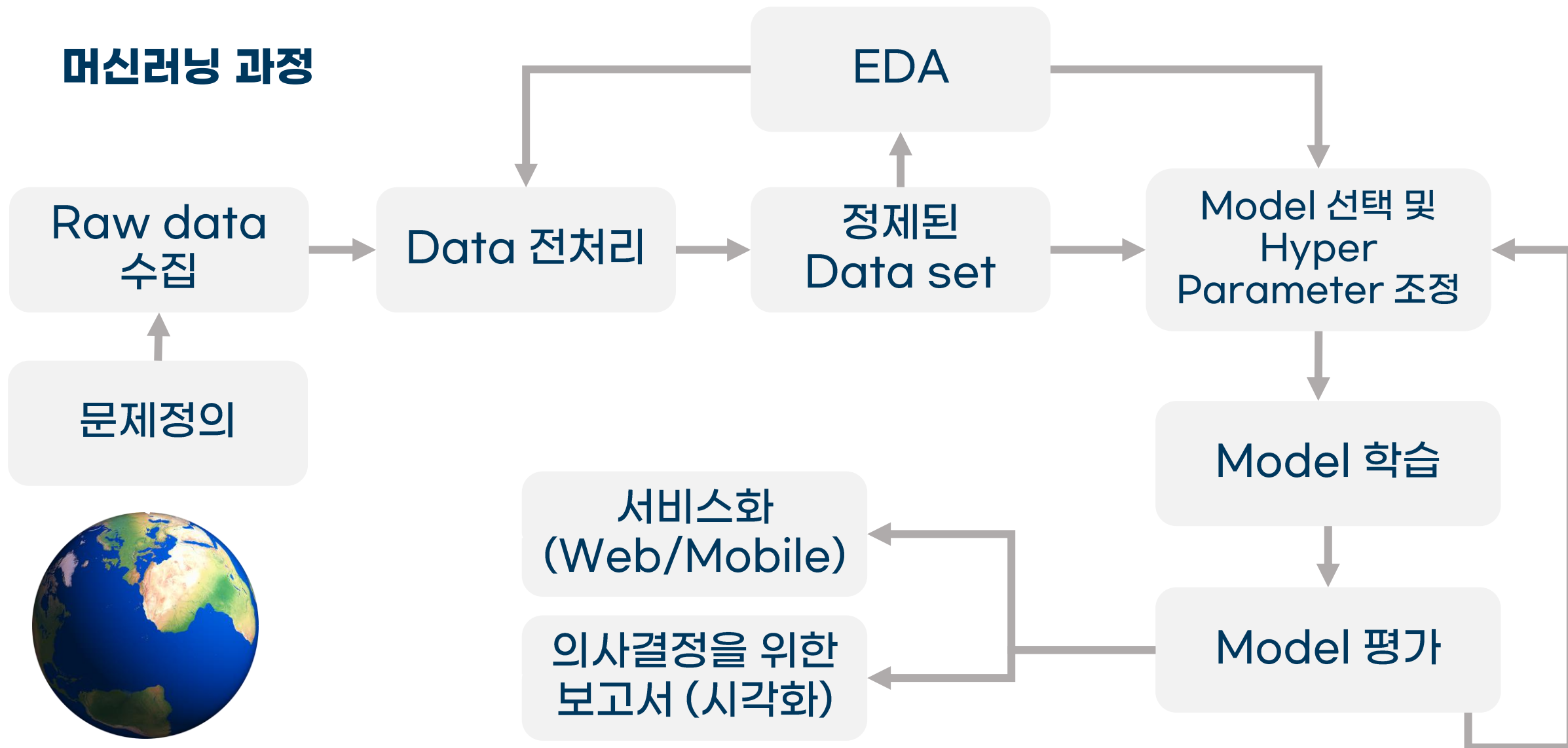
- accuracy(정확도)
- recall(재현율)
- precision(정밀도)
- f1 score
- roc곡선의 auc

### 회귀

- MSE(Mean Squared Error)
- RMSE(Root Mean Squared Error)
- $R^2$  (R Square)



## 머신러닝 과정



## AND 연산을 학습해 보자

문제[Data]

답[Label]

A	B	A and B	A or B	A xor B
0	0	0	0	0
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

## 머신러닝 모델 학습을 위한 도구

### scikit-learn

- 파이썬에서 쉽게 사용할 수 있는 머신러닝 프레임워크, 라이브러리
- 회귀, 분류, 군집, 차원 축소, 특성 공학, 전처리, 교차 검증, 파이프라인 등 머신러닝에 필요한 기능을 갖추
- 학습을 위한 샘플 데이터도 제공

## AND 연산 학습하기

Model  
(알고리즘)

```
knn_model=KNeighborsClassifier(n_neighbors=1)  
knn_model.fit(문제 , 답)  
knn_model.predict(예측값을 얻고 싶은 데이터 )  
score=metrics.accuracy_score(실제답, 예측결과 )
```

**KNN** 분류 모델 사용

## 실습 | 비만도 데이터 이용 학습해 보자

	A	B	C	D
1	Gender	Height	Weight	Label
2	Male	174	96	Obesity
3	Male	189	87	Normal
4	Female	185	110	Obesity
5	Female	195	104	Overweight
6	Male	149	61	Overweight
7	Male	189	104	Overweight
8	Male	147	92	Extreme Obesity
9	Male	154	111	Extreme Obesity
10	Male	174	90	Overweight
11	Female	169	103	Obesity
12	Male	195	81	Normal
13	Female	159	80	Obesity
14	Female	192	101	Overweight
15	Male	155	51	Normal
16	Male	191	79	Normal
17	Female	153	107	Extreme Obesity
18	Female	157	110	Extreme Obesity
19	Male	140	129	Extreme Obesity
20	Male	144	145	Extreme Obesity

회귀 문제?  
분류 문제?

# 다음 시간에 만나요!

일반화, 과대, 과소, KNN →