



모든 교사를 위한 **인천형**  
**인공지능교육**

① 이론 및 실습 편

모든 교사를 위한 **인천형**

# 인공지능교육

—— 이론 및 실습편

# 서문

“인공지능은 인류에게 작동하고 있는 가장 심오한 것 중 하나이다.

불이나 전기보다 더 심오하다.”

- 순다르 피차이 (2020년) 구글 최고경영자.

인류의 최고의 발견이자 발명품인 불과 전기보다 더 심오하다는 평가를 받는 인공지능. 인공지능은 인류의 삶을 어디까지 바꿀 수 있을까? 불과 출시 2달만에 월 사용자수(MAU) 1억을 달성한 인공지능(AI) 챗봇 ‘챗GPT’(ChatGPT)는 우리 삶에 어느새 깊숙이 파고든 인공지능의 존재를 강렬하게 인식하기에 부족함이 없다.

각종 기술과 도구가 인간의 삶에서 제대로 활용되고 올바로 쓰이기 위해서는 교사의 역할이 크다는 것에 이견이 없을 것이다. 이에 교사가 먼저 읽고 배워 학생들을 가르치자라는 취지로 개발된 것이 바로 ‘모든 교사를 위한 인천형 인공지능교육’이다.

‘모든 교사를 위한 인천형 인공지능교육’은 교육의 주인공 중 교사를 위한 교재이다. ‘모든 교사’는 초, 중등 교사와 모든 교과 교사들, 또한 인공지능이 무엇일까? 어떻게 가르칠까? 왜 가르칠까?에 대한 근본적인 의문을 가진 교사를 모두 아우르는 개념이다.

‘모든 교사를 위한 인천형 인공지능교육’은 2022 개정교육과정에 발맞춰 교육과정을 정리하고 인공지능의 기본 이해부터 최근 화두로 떠오른 ChatGPT의 다양한 활용 방법까지 충실히 신도록 노력하였다. 행간에 모두 담지 못한 내용은 부록인 PPT에 담았고, 교재 활용연수 동영상을 통해 이해도를 높일 수 있도록 연계하였다.

디지털 대전환의 시대,

모든 선생님들이 변화의 큰 파도에 휩쓸리기 보다 그 파도에 당당히 올라서는 ‘서퍼’가 되길 바래본다.

모든 교사를 위한 인천형 인공지능교육 집필진 일동

모든 교사를 위한 인천형  
인공지능교육

(중등 교사용 가이드북 - 이론 및 실습편)



## I. AI에 대한 이해

01. 컴퓨터는 어떻게 동작할까?	10
02. 인공지능이 뭐지?	17
03. 인공지능의 테크닉스는?	22
04. 인공지능 테크놀로지가 뭐지?	35

## II. AI와 데이터, 윤리

01. 데이터가 왜 중요하지?	48
02. 데이터를 어떻게 다루지?	55
03. 윤리 문제의 종류와 발생 원인은?	61
04. 윤리 문제 해결을 위한 교육, 그리고 법과 사회 제도	67



## I. AI로 데이터 시각화 분석

★ 엔트리란?	76
01. AI로 데이터를 시각화 및 분석해 보자(엔트리)	79
★ 오렌지3란?	89
02. AI로 데이터를 시각화 및 분석해 보자(오렌지3)	90

## II. AI로 응용프로그램 개발

01. ML2 스크래치	106
02. 핸즈포즈	118

## III. AI로 디바이스 제어

01. 티쳐블머신 웹페이지 제어	127
02. 티쳐블머신 아두이노 제어	133



AI 이론



AI 앱 제작실습



AI 활용

# I. AI에 대한 이해



이 단원에서는 인공지능에 대한 이론적인 이해를 도모합니다. 컴퓨터에 대한 작동원리를 이해하면 컴퓨터가 세상을 이해하는 방식과 컴퓨터로 어떤 문제를 해결할 수 있는지 알게 됩니다. 인공지능도 결국 컴퓨터의 물리적 기반에서 작동하기 때문에 컴퓨터가 어떤 원리로 다양한 일을 처리하는지 알게 되면 인공지능에 대한 통찰을 쉽게 얻을 수 있게 될 것입니다. 따라서 컴퓨터의 원리에 대해 배우고, 컴퓨터를 기반으로 한 인공지능 프로그램의 원리를 익힙니다. 그리고 인공지능을 구현하게 만들어주는 알고리즘과 그로 인한 테크놀로지들에 대해 살펴보도록 하겠습니다.

1. 컴퓨터는 어떻게 동작할까?
2. 인공지능이 뭐지?
3. 인공지능 테크닉스
4. 인공지능 테크놀로지

## 01

# 컴퓨터는 어떻게 동작할까?

생각  
열기

- A: 컴퓨터는 왜 0과 1밖에 모르는 거야?  
 B: 전기를 먹고 살아서 그런 것 아닐까? 안통하면 0, 통하면 1  
 A : 진짜 그런 것 같은데?

## 1. 컴퓨터의 탄생

애니악



에드삭



### ① 컴퓨터의 개념

계산 가능한 기계에 대한 개념을 최초로 만든 사람은 앤런 튜링(1912~1954)이다. 1945년에 고안한 초보적인 컴퓨터 튜링머신은 복잡한 계산과 논리 문제를 처리할 수 있었다. 계산 가능 기계의 지능을 판별할 수 있는 튜링 테스트도 고안하여 ‘컴퓨터과학의 아버지’라고도 불린다.



### ② 컴퓨터의 구현

계산할 때마다 수동으로 진공관의 회로 스위치를 다시 조정하여 새 입력을 처리하는 외부 프로그램 입력 방식의 애니악이라는 컴퓨터가 주로 사



용되었다. 이러한 방식은 준비시간이 너무 오래 걸리고, 전력소모가 심했다.

폰 노이만이 프로그램 내장방식 컴퓨터를 최초로 제안하고, 이를 바탕으로 전자식 컴퓨터 에드삭이 만들어졌다. 폰 노이만 구조는 중앙처리 장치, 메모리, 프로그램 세 가지 요소로 구성되어 있다. 현대 모든 컴퓨터의 구조는 폰 노이만 구조를 기반으로 한다.

노이만 구조 컴퓨터는 프로그램 입력, 처리, 결과 출력으로 구분된다. 처리 장치에는 제어장치, 산술/논리장치, 각종 고속기억장치인 레지스터로 구성되고, 명령과 데이터를 저장할 메모리로 구성된다.

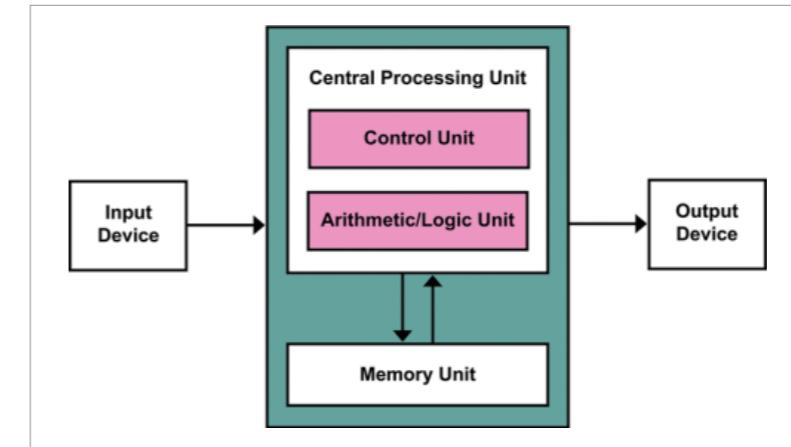


그림 1-3\_폰 노이만 구조

## 2. 컴퓨터 하드웨어의 원리

### ① 물리적 구성요소

현대 컴퓨터의 핵심 처리 장치 CPU는 반도체인 트랜지스터의 조합으로 만든다. 트랜지스터는 스위치 역할을 하는데 사람이 전기를 이용해 ON/OFF를 마음대로 조종할 수 있다. 컴퓨터는 전기가 통하면 1, 전기가 안통하면 0으로 추상화한 2진수를 사용한다. 2진수를 이용해 덧셈, 뺄셈, 곱셈, 나눗셈, 논리연산, 저장 등을 할 수 있게 만들면 컴퓨터가 만들어 진다.

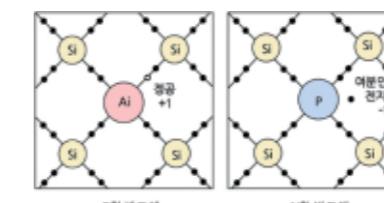
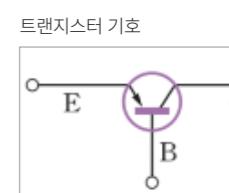


그림 1-4\_트랜지스터의 원리



## AND연산과 게이트

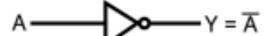
A	B	$A \times B$
0	0	0
0	1	0
1	0	0
1	1	1

## OR연산과 게이트

A	B	$A + B$
0	0	0
0	1	1
1	0	1
1	1	1

## NOT연산과 게이트

A	$\bar{A}$
0	1
1	0



논리연산에는 AND, OR, NOT연산이 있으며, 이 연산을 수행할 수 있는 장치를 게이트라한다. 게이트를 조합하면 컴퓨터를 구현할 수 있다. 트랜지스터 2개를 직렬로 연결하면 AND연산을 할 수 있는 장치인 AND게이트를 만들 수 있다.

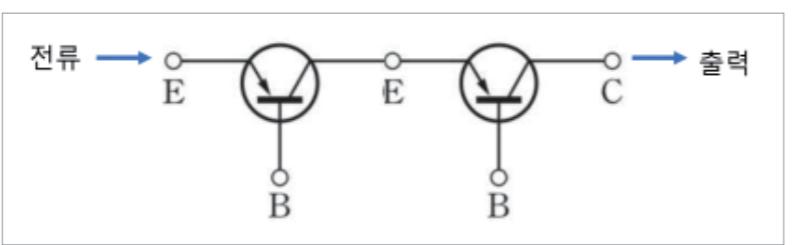


그림 1-5\_트랜지스터의 직렬 연결

트랜지스터 2개를 병렬로 연결하면 OR연산을 할 수 있는 장치인 OR 게이트를 만들 수 있다.

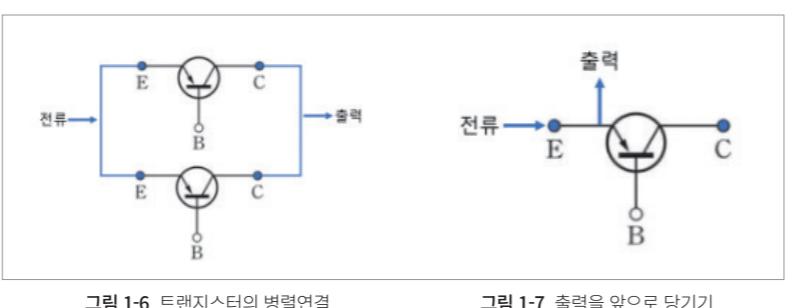


그림 1-7\_출력을 앞으로 당기기

트랜지스터의 출력을 앞으로 당기면 NOT연산을 할 수 있는 장치인 NOT게이트를 만들 수 있다.

## ② 2진수의 덧셈

2진수 각 자리값은 1, 2, 4, 8로 2배씩 증가하고, 크기를 0과 1로 나타낸다. 크기가 0인 자리값은 제외하고, 1인 경우의 자리값만 더하면 10진수로 쉽게 변환된다. 예를 들면 1001은 1인 자리값만 더하면 9가 된다. 한 자리의 2진수 2개를 더할 수 있는 장치를 만들 수 있다면 더 많은 자리의 2진수를 더할 수 있을 것이다. 2진수의 덧셈은 논리연산을 이용해 구현할 수 있다.



그림 1-8\_2진수 체계

입력 A와 B에 2진수 한자리씩 입력된 것을 계산하면 합과 올림수가 출력으로 나오게 되고, A와 B의 경우의 수에 따라 합의 진리표와 올림수의 진리표가 나오게 된다.

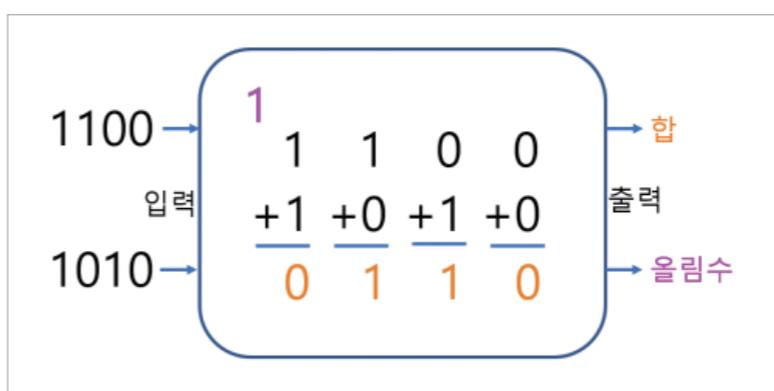


그림 1-9\_2진수의 덧셈

올림수의 진리표는 AND연산의 결과와 같아 AND게이트를 사용하면 구현할 수 있다.

합의 진리표는 X-OR연산의 결과와 같고, XOR연산은 아래와 같이 AND, OR, NOT게이트를 조합하면 구현할 수 있다.

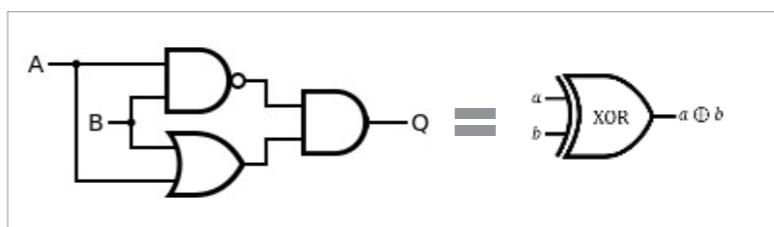


그림 1-10\_XOR연산의 구현

그림 1-11\_XOR게이트

AND게이트 2개, OR게이트 1개, NOT게이트 1개를 조합하면 합의 진리표와 같은 결과를 출력하는 XOR연산을 구현할 수 있게 된다. 게이트 4개가 조합된 게이트는 XOR게이트 하나로 나타낼 수 있다.

## ③ 덧셈기의 구현

한 자리 2진수 A와 B를 더할 수 있는 덧셈기는 다음과 같이 XOR게이트와 AND게이트를 이용해 구현이 가능하다.

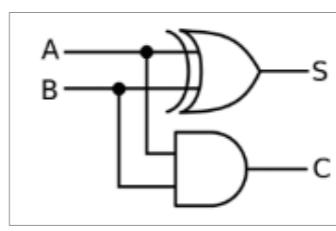


그림 1-12\_반가산기

## 합의 진리표

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

## 올림수의 진리표

A	B	S	C
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

## XOR연산과 게이트

입력		출력	
A	B	Y	
0	0	0	
0	1	1	
1	0	1	
1	1	0	

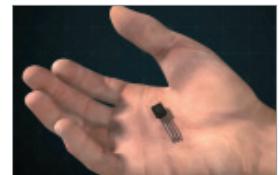
  

## 반가산기

2진수의 한 자리 수를 더할 수 있고, 올림수가 발생한다.

전가산기  
2진수의 한 자리 수를 더할 수 있고, 아래 올림수까지 반영할 수 있다

트랜지스터의 크기



레지스터

CPU가 요청을 처리하는데 필요한 데이터를 일시적으로 저장하는 다목적 공간

두 번째 자리 이상의 2진수 A와 B, 그리고 아래에서 올라온 올림수를 더할 수 있는 덧셈기는 다음과 같이 XOR게이트와 AND게이트, OR게이트를 이용해 구현이 가능하다. 실제로 모든 수를 계산할 수 있는 덧셈기는 이것보다 복잡하다.

많은 트랜지스터를 조합하여 게이트를 만들고, 게이트를 조합하여 덧셈과 논리연산이 가능한 산술논리 장치(ALU)를 만들어 2진수의 계산이 물리적으로 구현 가능해졌다.

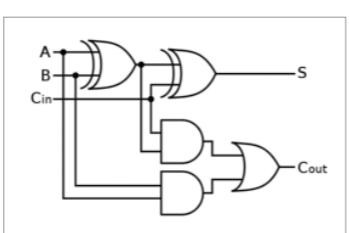


그림 1-13\_전가산기

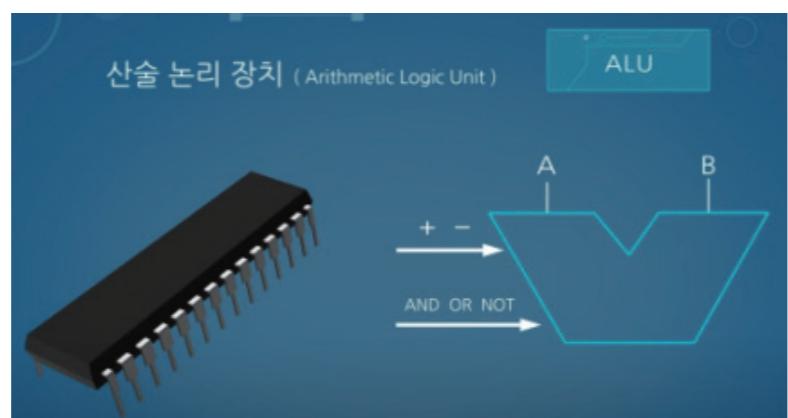


그림 1-14\_산술논리장치

이론적으로 빛의 속도로 계산이 가능한 덧셈기를 이용해 덧셈을 여러번 수행하여 곱셈을하고, 보수로 만들어 더함으로해서 뺄셈이 가능하고, 뺄셈을 이용해 나눗셈을 하는 사칙연산이 가능하다. 컴퓨터를 우리말로 전자계산기라하는 이유가 여기에 있다. 컴퓨터는 정말로 0과 1밖에 모르지만 트랜지스터의 스위치 기능을 이용해 사칙연산, 논리연산과, 레지스터를 이용한 저장 등의 기능을 수행해 다양한 일을 할 수 있다. 결국 트랜지스터로 이루어진 게이트로 논리연산과 산술연산이 가능한 전자계산기(컴퓨터)를 만들어 사용할수 있게 되었다.

### 3. 컴퓨터 프로그램 동작의 원리

#### ① 프로그램의 번역과 실행

지금까지 살펴본 바와 같이 컴퓨터는 2진수만 처리할 수 있고, 이해한다. 하지만 사람은 무수히 많은 2진수의 나열을 보고 해석하거나 2진수로 컴퓨터에게 명령할 수 없다. 따라서 인간이 사용하는 자연어에 가까운 고급언어를 이용해 컴퓨터에게 수행시킬 명령어를 작성하고, 이를 번역해 2진수로 바꿔 줄 번역기가 필요하다. C언어로 작성된 프로그램은 컴파일러에 의해 어셈블리 언어로 번역하고 다시 어셈블러에 의해 기계어로 번역되어 명령과 주소, 데이터로 이루어진 2진수의 조합으로 바뀌게 된다.

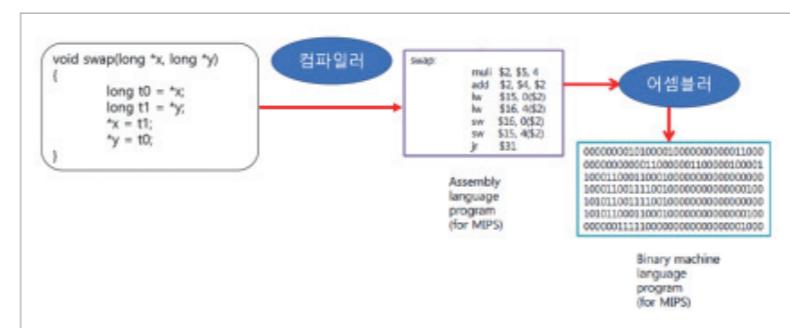


그림 1-15\_코드의 번역 과정

$C = A + B$ 를 컴파일러가 어셈블리언어인 LOAD[10]으로 번역하고, 어셈블러가 LOAD[10]을 기계어인 '1001000000001010'으로 번역한다. 번역된 2진수는 인출, 해석, 실행, 저장의 순서로 순차적으로 실행된다.



그림 1-14\_산술논리장치

예를 들어  $C = A + B$ 라는 단순한 덧셈 계산을 수행할 때 10번지에 있는 값 2를 로드하고, 11번지에 있는 값 3을 더하여 12번지에 결과 5를 저장하라는 명령이 순차적으로 메모리와 연동하여 수행된다.

컴파일러  
작성한 코드를 한꺼번에 번역하는 프로그래밍언어 번역기

어셈블러  
어셈블리 언어로 작성 된 코드를 기계어로 번역하는 번역기

인출-해석-실행-저장  
인출: 명령어를 메모리에서 가져오는 과정  
해석: 명령어의 의미를 해석하는 과정  
실행: 명령을 수행하는 과정  
저장: 결과를 메모리에 저장하는 과정

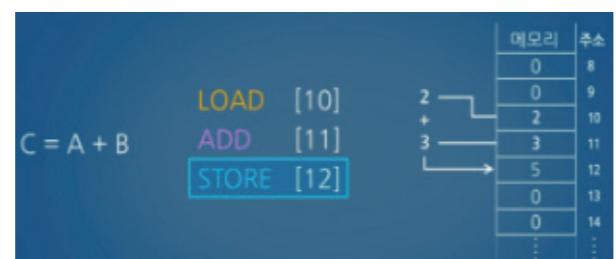


그림 1-17\_프로그램 실행 원리

## ② 알고리즘

컴퓨터 내부에서 명령어가 수행되는 원리에 비춰보면 이 과정이 잘 처리되기 위해서는 프로그래머가 논리적인 오류가 없도록 명령을 만들어 줘야 한다. 컴퓨터가 수행해야 할 문제에 대한 순서대로 나열한 지시사항을 알고리즘이라 한다. 고급언어인 C언어나 파이썬을 이용해 컴퓨터에게 시킬 명령을 작성하기 전에 논리적이며 수행 가능하고, 언어 문법에 맞는 알고리즘을 만들어 줘야 한다. 알고리즘을 잘 작성하면 오류 없이 번역이 되고, 컴퓨터가 수행할 수 있게 된다.

알고리즘을 구상하여 컴퓨터에게 수행할 시킬 일의 순서를 만드는 사람을 프로그래머라하며, 프로그래밍 언어를 도구로 프로그래밍 활동을 통해 명령의 집합체인 프로그램이 결과로 나온다. 이 프로그램은 번역기에 의해 2진수로 번역되어 하드디스크에 저장되었다가 실행 시키면 메모리로 이동하고, 중요 명령과 데이터가 CPU의 레지스터에 저장되어 순차적으로 실행된다.

### 알고리즘

특정한 문제를 풀기 위해 거쳐야 하는 논리적인 일련의 과정들을 묶어놓은 절차나 방법

## ③ 계산 가능 문제

컴퓨터는 논리게이트를 기반으로 구동되는 2진 계산 장치이기 때문에 컴퓨터에 입력해 주는 프로그램은 계산 가능 해야 한다. 대부분의 컴퓨터 과학자들은 계산 가능 문제에 집중해 왔다. 수학의 함수처럼 입력 X가 들어오면 함수 내부의 수식을 만들어 계산 가능한 알고리즘을 연구하여 출력 Y가 나와야 하는 것이다. 우리가 사용하는 대부분의 프로그램이 모두 이런 원리로 만들어진다.

일상의 문제들은 정확한 정답이 있는 문제 보다는 대종, 어림짐작, 둘 중의 하나 등으로 해결 가능한 문제가 더 많다. 보통 이런 문제를 휴리스틱 문제라 하는데 이런 문제는 컴퓨터로 해결할 수 없었다. 예를 들어 개와 고양이 사진을 보고 개인지 고양이인지 판별하는 알고리즘을 만들기는 인간의 능력으로 거의 불가능하였다. 이러한 휴리스틱 문제는 인공지능 SW가 더 잘 해결한다.

### 휴리스틱

시간이나 정보가 불충분하여 합리적인 판단을 할 수 없거나, 굳이 체계적이고 합리적인 판단을 할 필요가 없는 상황에서 신속하게 사용하는 어림 짐작의 기술

# 02

## 인공지능이 뭐지?

생각  
열기

A: 지능을 인공적으로 만들었다는 거잖아?

B: 그렇지. 컴퓨터로 지능을 구현했다는 말이지

A: 그럼 인공지능도 0과 1밖에 모르는 컴퓨터로 구현한 거네?

## 1. 인공지능의 개념

### ① 인공지능의 정의

인공지능(AI, artificial intelligence) 이란 **인간의 지능**을 컴퓨팅 기계를 통해 인공적으로 재현하는 기술을 의미한다. 즉, 인공지능이 적용된 소프트웨어나 제품들은 프로그램된 처리를 넘어 인간이 지닌 지능과 비슷한 지능적인 처리를 수행할 수 있다. 인간의 지능은 다양한 지능의 복합 지능이지만 인공지능은 인간의 지능 중에서 인식, 학습, 추론, 분류, 예측, 판단, 생성 같은 일부 지능만 구현한 것이다. 현재는 일반 SW와 인공지능 SW로 구분하지만 시간이 지나면 모

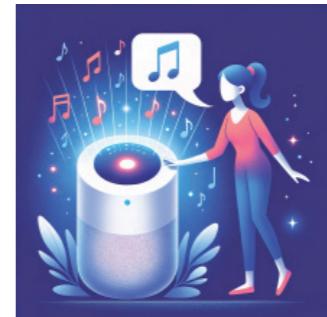


그림 2-1\_음성인식 인공지능

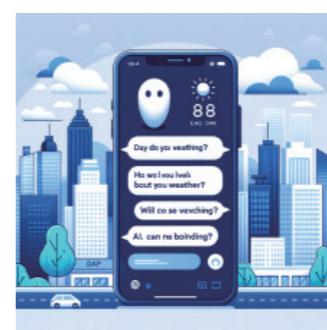


그림 2-2\_인공지능 챗봇

든 SW나 컴퓨터 프로그램은 인공지능 기능이 일부 또는 전부 탑재되어 사실상 구분이 의미 없게 될 것이다.

## ② 인공지능의 특성

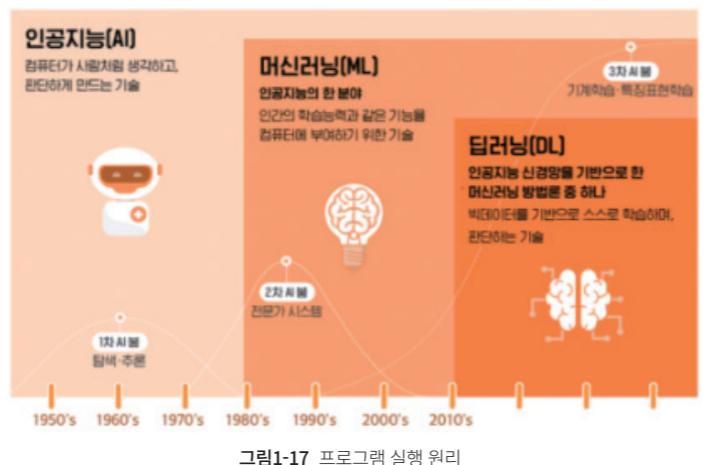
- 가) 인식 : 인간은 감각기관을 통해 다양한 자극을 감지하고 인식하는데, 인공지능은 센서를 통해 입력된 데이터를 해석하여 인식한다. 인공지능은 음성 인식, 이미지 인식이 가능하다.
- 나) 학습 : 인간은 경험을 통해 학습하는데, 인공지능은 주어진 많은 데이터에서 정답이 나오는 패턴을 찾아내 학습한다.
- 다) 추론 : 인공지능은 규칙, 논리 등으로 표현된 지식을 바탕으로 새로운 지식을 빠르게 추론한다.
- 라) 생성 : 인공지능은 학습을 통해 앞의 단어를 보고 다음에 나올 단어를 계산을 통해 최대의 확률을 가진 단어를 생성해 낸다.

## ③ 인공지능 알고리즘

인공지능 SW도 역시 폰 노이만 구조의 컴퓨터에 기반해 트랜지스터로 이루어진 0과 1만 사용하는 컴퓨터에서 구동된다. 일반 프로그램이 어떤 값을 입력하면, 작성된 프로그램의 내용에 따라 정해진 응답을 출력하는 것과 동일한 물리적인 시스템에서 실행되는 것이다. 따라서 인공지능도 가장 단순한 0과 1의 계산에 의해 지능적인 출력 결과를 도출한다. 초기의 인공지능은 인간이 만든 규칙기반 시스템이나 탐색알고리즘, 최적화 알고리즘 등에 기반하여 작동되었다. 얼마 지나지 않아 논리적인 규칙기반의 알고리즘을 정교하게 만드는 데는 한계가 있다는 것을 알게 되었고 딥러닝 기술이 나오기 전까지 몇 번의 상승과 하강 곡선을 그리며 발전하였다.

### 인공지능 암흑기

1, 2차 AI붐 이후에 나타난 시기로 더 이상 발전이 없이 인기가 사그라든 시기



## 2. 머신러닝

### ① 1차 방정식

일반 SW를 가장 단순한 수식으로 다음과 같이 1차 방정식으로 나타낼 수 있다.

$$y = ax + b$$

사용자가 입력  $x$ 를 넣으면 출력  $y$ 를 얻게 되는데 우리가 사용하는 모든 SW는 수학적으로 함수라 할 수 있다.

$$f(x) = ax + b$$

여기서 SW를 만드는 프로그래머나 컴퓨터 과학자는 함수의 처리 내용에 해당하는  $a$ 와  $b$ 를 고민하고, 결정하는 역할을 한다. 복잡한 프로그램일 경우 함수식을 결정하는 일은 매우 어려운 일이다. 그래서 SW를 만드는 일은 전문적인 영역의 일이었다.

생각을 뒤집어  $a$ 와  $b$ 를 결정하는 일을 컴퓨터에게 맡긴다면 전문적인 영역이었던 프로그램을 개발하는 일을 컴퓨터에게 맡길 수 있을 것이다.

전통적인 프로그래밍에서는 문제의 해결 방법을 코드로 직접 작성해야 했으나 머신러닝에서는 데이터를 통해 ' $a$ '와 ' $b$ '와 같은 파라미터 값을 자동으로 조절하며, 최적의 함수를 학습하게 된다. 이렇게 학습된 함수는 다양한 입력  $x$ 에 대한 예측값  $y$ 를 제공하며, 이 과정을 통해 데이터의 숨겨진 패턴이나 관계를 발견하게 된다. 따라서, 프로그래머는 더 이상 모든 상황에 대한 규칙을 정의하는 대신, 적절한 학습 알고리즘과 데이터를 사용하여 컴퓨터가 스스로 규칙을 학습하게 만드는 역할을하게 된다. 컴퓨터는 많은 데이터로부터 함수식을 결정하고, 결정된 함수식은 과거 프로그래머가 고민 끝에 만들어낸 함수식과 같은 것이다. 사용자는 함수식에 입력을 넣으면 결과가 나오게 되는 것이다.

**파라미터**  
프로그램 상에서 조절이 필요한 변수

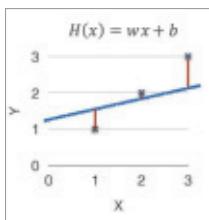
**가설식**  
 $w = \text{Weight}$   
 $b = \text{Bias}$

### ② 인공지능 학습의 과정

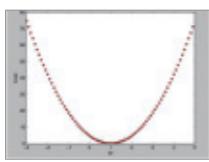
통계학에서 사용하는 선형회귀식을 이용해 가설식으로 사용한다.

$$f(x) = ax + b \rightarrow H(x) = wx + b$$

cost  
가설H와 실제 데이터간의 차이  
 $H(x)-y$



cost가 0이 되는 지점 찾기  
W-cost = 0



임의의 w와 b값을 이용해 임의의 기울기의 직선을 그는다. 여기서 'w'와 'b'는 파라미터이다. 입력 x와 결과 y의 데이터를 바탕으로 이 'w'와 'b' 값을 조절(학습)하여 가장 데이터를 잘 나타내는 직선을 찾아낸다. 여기서 정답 데이터 쌍인 x, y값은 주어야 하며, 아주 많은 데이터 쌍이 있어야 파라미터를 조절할 수 있게 된다. 임의로 잡은 가설식과 정답 y의 차이를 모두 더한 것을 cost라고 하며, cost가 0이 되어야 가설식은 정답에 가까운 것이 되므로 cost가 0이 되는 지점을 찾는 것이다. 이러한 과정을 우리는 기계가 학습한다하여 머신러닝이라 부른다.

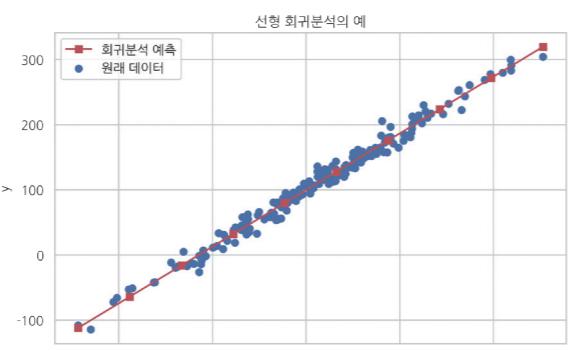


그림 2-4\_선형회귀분석 그래프,

변수가 여러개인 경우  
짜장/짬뽕 선택에 있어서도 날씨, 취향, 타인의 선택 등 다양한 요인이 결정에 영향을 미치게 되므로 변수가 많이 필요하다.

우리 주변에는 짜장/짬뽕, 개/고양이 등 선택에 대한 문제가 많다. 이러한 문제는 가설식  $H(x)$ 가 정답으로 주어진 데이터상으로부터 cost가 0이 되는 절편과 기울기를 결정하고 나면 함수식이 결정되고, 학습이 종료되었다고 한다. 새롭게 들어오는 X데이터가 어느 위치인지 결정이 되기 때문에 컴퓨터 입장에서 지능이 생긴 것과 같은 결과를 내놓게 된다. 여기서 결과 y에 영향을 끼치는 요인이 많을 경우 변수를 하나만 사용하지 않고 여러개를 사용하여 구할 수 있다.

$$y = w_1x_1 + w_2x_2 + w_3x_3 + b$$

알맞지 않다. 인간 뉴런을 모델링하여 수식으로 나타내면 여러 개의 입력이 존재하는 1차 방정식과 같다.

$$y = w_1x_1 + w_2x_2 + w_3x_3 + b$$

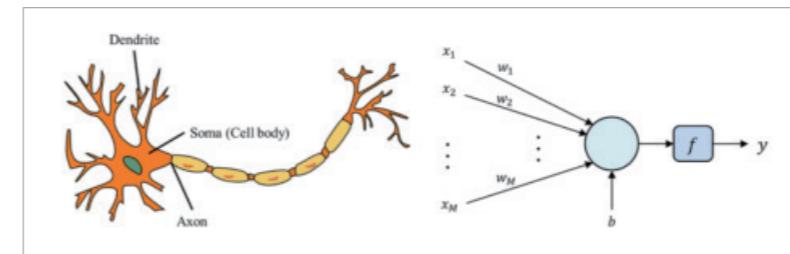


그림 2-5\_생물체의 뉴런과 인공뉴런

인공뉴런(AN)은 퍼셉트론이라 불리며 딥러닝에 있어서 연산의 최소단위이다. 여러개의 입력이 서로 다른 가중치를 곱한뒤에 모두 더해 임의의 임계값보다 크면 1을 출력하고 그렇지 않으면 0을 출력한다. 퍼셉트론은 간단한 이진 분류문제를 풀기 위해 연구되었으나 하나만 가지고 어려운 문제를 해결하기는 힘들다. 따라서 퍼셉트론을 다층으로 쌓고, 레이어를 여러층으로 나누어 연결하면 인공신경망(ANN)이 구성되어 복잡한 문제를 해결할 수 있게 된다. 인공신경망이 여러 계층을 이루고 있어 딥러닝이라 부르며, 딥러닝의 대표적인 알고리즘은 다음과 같다.

- 가) 합성곱 신경망 (Convolutional Neural Networks, CNN) – 주로 이미지 처리에 사용
- 나) 순환 신경망 (Recurrent Neural Networks, RNN) – 주로 시퀀스 데이터 처리에 사용
- 다) 트랜스포머 (Transformer) 모델 – 주로 자연어 처리에 사용
- 라) Generative Adversarial Networks (GANs) – 데이터 생성에 사용

### 3. 딥러닝

#### ① 딥러닝의 이해

데이터를 통해 모델을 학습시키는 머신러닝 알고리즘은 분류, 예측 등에 사용이 가능하지만 자연어나 이미지같은 시각 정보를 처리하는데는

#### ② 일반 SW 인공지능 SW

결국 일반 SW는 프로그래머가 함수 내부의 수식을 작성한다면 인공지능SW는 함수 내부의 수식을 주어진 많은 데이터로부터 학습을 통해 컴퓨터가 결정한다. 학습 알고리즘과 빅데이터만 제공하면 데이터로부터 패턴을 찾아내 문제해결의 최적의 함수를 결정하게 된다. 일반SW나 인공지능SW 모두 0과 1의 계산에 의해 논리적인 절차나 데이터로부터 패턴을 찾는 일을 수행하는 것이다.

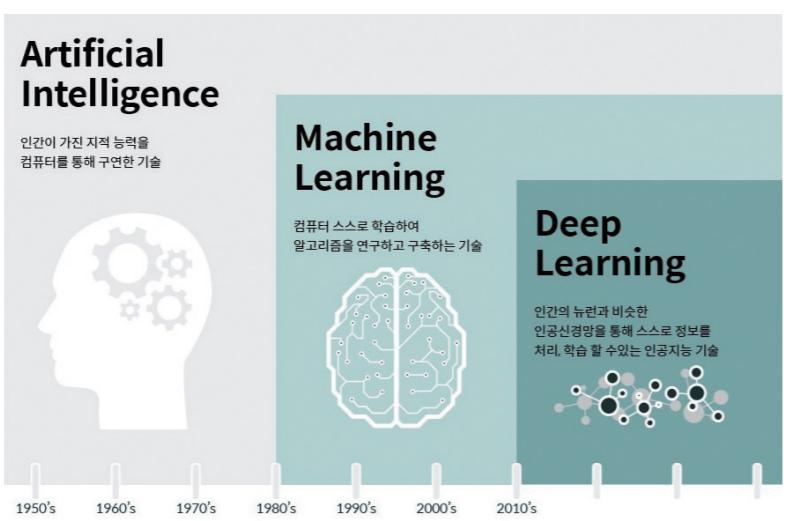
## 03

인공지능의  
테크닉스는?

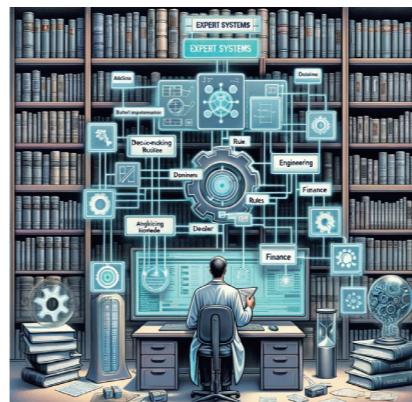
- Q. 인공지능과 기계학습 딥러닝의 관계는 무엇일까?  
 Q. 전문가 시스템이 더 발전하지 못한 이유는 무엇인가?

## 1. 기계학습

## ① 기계학습 개요



기계학습을 이해하기 위해서는 우선 인공지능과 기계학습, 딥러닝의 관계를 알아야한다. 인공지능은 인간의 지적능력을 컴퓨터로 구현하는 모든 기술과 학문을 총칭한다. 인공지능을 구현하기 위해 다양한 방법들이 연구되어 왔고 그 중의 하나가 기계학습이다. 기계학습은 단순히 인공지능을 구현하는 방법 중의 하나가 아니라 현재 인공지능의 기술발전의 근간이 되는 연구 분야이다. 딥러닝은 생물의 신경망 이론과 기계학습이 결합된 학습방법을 의미한다.



기계학습이 대중화되기 전 인공지능을 구현하는 대표적인 방법 중 하나로 규칙 기반 접근 방식인 전문가 시스템이 사용되었다. 전문가 시스템은 인간 전문가가 알고 있는 규칙들을 추상화 시켜 표현하고 그것을 컴퓨터에 입력하여 구현한다. 예를 들어 의료 분야 전문가 시스템을 만들기 위해

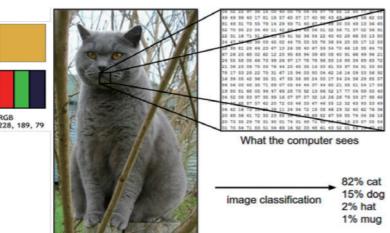
서는 의사가 환자의 상태를 보고 진단을 내릴 때 사용하는 판단 기준과 규칙들을 명시화하여 시스템에 입력한다. 많은 규칙을 넣을수록 시스템은 더 정확한 결과를 도출하게 된다. 하지만 이 전문가 시스템은 오래 가지 못했다. 그 이유는 무엇일까?

전문가 시스템을 구현할 때 모든 규칙을 입력할 수 없기 때문에 시스템에 입력되는 상황이 조금만 달라져도 올바른 해결책을 찾지 못한다. 또한 규칙 중에는 쉽게 추상화하여 표현할 수 없는 것들도 존재한다. 다음 예를 한번 보자.



사진 속 강아지와 고양이를 구분하는 문제를 살펴보자. 사람에게 두 객체를 구분하는 것은 쉽지만 그 기준을 말로 표현하기는 어렵다. 우리는 태어나서부터 강아지와 고양이를 실물과 사진, 영상 등으로 많이 봐왔기 때문에 둘의 미세하고 복잡한 차이(컴퓨터의 기준)를 자연스럽게 학습할 수 있었다. 사람은 많은 경험을 통해 패턴을 자연스럽게 학습했기 때문에 학습한 패

턴을 다시 명시적으로 표현하는데 어려움을 느낀다. 고양이와 강아지를 구분하는 것도 이렇게 표현하기 어려운데 다른 일들은 어떨까? 세상이 발전하면서 컴퓨터로 해결해야 하는 문제들은 더 복잡하고 미묘해져가고 있다. 만약 우리가 세상에 있는 복잡하고 미묘한 규칙들을 일일이 추상화하여 표현해야한다면 시간과 비용이 얼마나 들까? 상상도 못할 정도다.

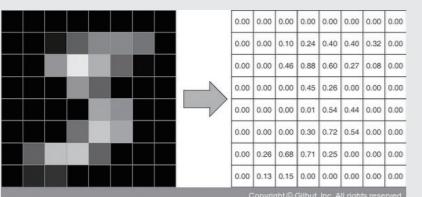


그래서 등장한 것이 기계학습이다. 기계학습은 사람이 규칙을 직접 입력하는 것이 아니라, 컴퓨터가 스스로 데이터를 통해 복잡한 규칙을 찾아내고 학습하는 방법이다. 이것은 마치 우리가 고양이와 강아지를 구분하는 능력을 학습하는 것과 유사하다. 이로써 우리는 복잡한 규칙을 어렵게 추상화할 필요도 없고 모든 규칙을 일일이 입력해 줄 필요도 없게 된 것이다. 이어지는 파트에서 어떻게 컴퓨터가 인간처럼 학습하는지 알아보도록 한다.



더 알아보기

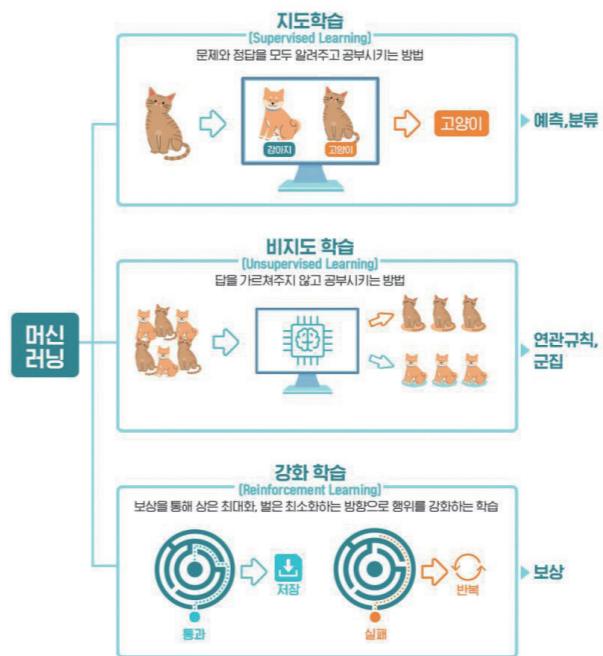
#### 이미지의 디지털 표현방식 - 비트맵(bitmap) 이미지:



비트맵 이미지는 픽셀이라는 작은 점들의 그리드로 구성되어 있다. 각 픽셀은 특정한 색상 값을 가지며, 이러한 픽셀들의 집합이 전체 이미지를 형성한다. 색상 값은 보통 RGB(빨강, 초록, 파랑) 색상 모델을 사용하여 표현되며, 각 색상은 0부터 255까지의 값으로 표현된다. 해상도는 이미지의 세부 사항을 결정하며, DPI(Dots Per Inch) 또는 PPI(Pixels Per Inch)로 측정된다. JPEG, PNG, BMP, GIF 등이 비트맵 이미지 형식의 예이다.

## ② 기계학습의 정의 및 종류

기계학습은 컴퓨터에게 데이터를 통해 스스로 학습하게 하여, 새로운 데이터나 상황에 대한 예측이나 결정을 자동으로 할 수 있게 하는 알고리즘과 기술의 집합이다. 단순히 말하면, 기계학습은 컴퓨터가 경험을 통해 스스로 문제해결 규칙을 학습하는 것을 의미한다. 기계학습에는



여러 방법들이 있는데, 그 중 대표적인 세 가지 유형을 살펴본다.

우선 학습 데이터의 정답(레이블) 유무에 따라 지도학습과 비지도 학습이 다르게 사용된다. 자세한 설명은 다음 파트에서 알아본다. 강화 학습은 에이전트가 환경과 상호 작용하며 보상과 처벌을 통해 최적의 전략을 학습하는 방식이다. 예를 들어 게임에서 로봇이 동전을 모으는 미션을 수행하게 한다고 할 때, 로봇이 동전을 잘 모으면 보상을, 잘못된 방향으로 움직이면 처벌을 받게 된다. 이렇게 반복하면 로봇은 동전을 효과적으로 모으는 방법을 스스로 학습하게 된다. 우리가 잘 알고 있는 알파고는 지도학습 방식과 강화학습 방식을 함께 사용하여 만들어졌다.



더 알아보기



기계학습 알고리즘도 프로그램으로 작성되어 실행됩니다. 그렇다면 기계학습 프로그램과 일반 프로그램의 차이점은 무엇일까요? 일반 프로그램은 문제해결 규칙을 알고리즘으로 표현하여 사람이 직접 구현하는 것입니다. 예를 들어 강아지 고양이를 구분하는 방법과 순서를 프로그래밍 언어로 직접 작성하는 것입니다. 기계학습은 데이터를 학습해 데이터 속 규칙을 찾아내는 것입니다. 예를 들어 고양이와 강아지 데이터를 주면 데이터에서 구분하는 방법과 순서를 찾아서 도출합니다. 이를 그림으로 표현하면 다음과 같습니다.

그럼 기계학습 방식 중 지도학습과 비지도 학습을 비유를 통해 쉽게 알아보도록 하자.

#### 지도학습 알고리즘 예:

선형 회귀, 로지스틱 회귀, 의사결정 트리, 랜덤 포레스트, 서포트 벡터 머신, 신경망

#### 비지도학습 알고리즘 예:

군집화, K-평균(K-Means), 계층적 클러스터링, 차원 축소, 주성분 분석, t-SNE, 연관 규칙 학습, Apriori, Eclat



지도학습은 선생님이 있는 아이가 학습하는 과정과 유사하다. 위 그림에서 아이는 선생님과 함께 사과와 농구공을 분류하는 방법을 학습한다. 이 때 선생님이 수많은 그림을 보여주면서 '이것은 사과고 저것은 농구공이야'라고 알려준다. 이렇게 선생님이 각 물건이 무엇인지 알려주면서 학습하게 되면 이 아이는 사과와 농구공을 구분할 수 있을 뿐만 아니라 내가 구분한 것이 무엇인지(사과, 농구공)도 알게 된다.



#### 알고리즘의 예:

지도학습에는 선형 회귀, 로지스틱 회귀, 결정 트리, 랜덤 포레스트, 서포트 벡터 머신 등이 있고, 비지도 학습에는 k-평균 클러스터링 계층적 클러스터링 주성분 분석(PCA), t-SNE 등이 있습니다.

비지도 학습은 독학하는 아이와 비슷하다. 아이는 선생님이 없어서 앞에 있는 동그란 물건이 무엇인지는 모르지만 수많은 구형을 물고 빨아보면서 먹을 수 있는 것과 없는 것으로 종류를 구분할 수 있게 된다. 결과적으로 다른 환경의 두 아이는 사과와 농구공을 구별할 수 있게 됩니다. 차이점이라면 지도학습한 아이는 그 물건이 무엇인지 알고 비지도 학습한 아이는 모른다는 것뿐입니다.

	지도학습	비지도학습
설명	지도 학습은 학습 데이터에 정답(레이블)이 포함. 컴퓨터에게 입력과 해당하는 정답을 함께 제시하며, 컴퓨터는 이 정보를 바탕으로 학습.	비지도 학습은 레이블이 없는 데이터를 사용. 컴퓨터는 데이터의 구조나 패턴을 찾아내려고 노력.
예시	강아지와 고양이 사진을 학습시키는 경우. 각 사진마다 '강아지' 또는 '고양이'라는 레이블을 붙여서 학습시키면, 새로운 사진이 주어졌을 때 그것이 강아지인지 고양이인지를 구분	쇼핑몰에서 고객들의 구매 이력만 바탕으로, 서로 비슷한 구매 패턴을 가진 고객들의 그룹(클러스터)을 만들어내는 경우.



Q. 지도학습과 비지도 학습 중 무엇이 더 효과가 좋을 것 같은지  
추측해보고 그 이유를 써보자.

## 2. 기계학습의 동작원리

### ① 지도학습의 동작원리

지도학습의 동작원리를 예시를 통해 자세히 살펴보자.

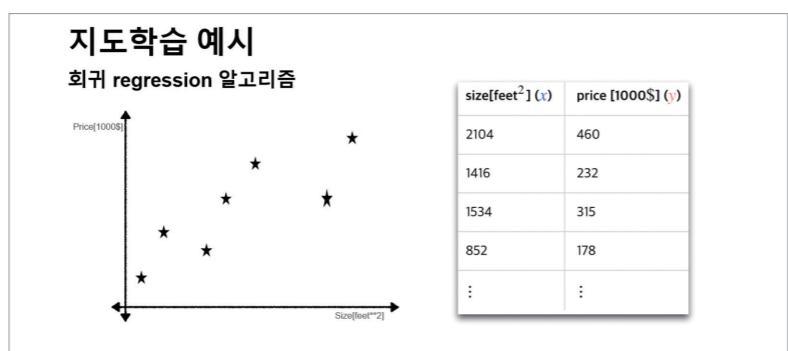


#### 첫 번째 예시(회귀)

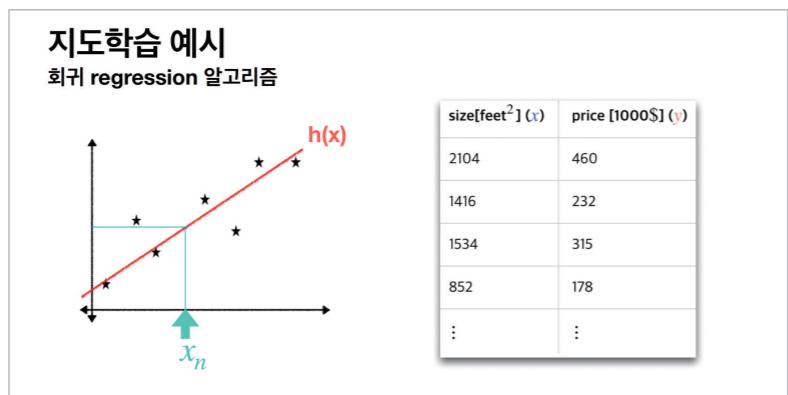
정국이는 미국으로 이사를 갈 예정이다. 그런데 미국살이는 처음이라

회귀 알고리즘은 통계학에서 유래한 머신러닝의 한 방법으로, 주어진 독립변수들과 종속변수 사이의 관계를 모델링하는데 사용됩니다. 목적은 하나 또는 그 이상의 독립변수에 대해 연속적인 값을 예측하는 것입니다.

집값이 얼마인지 잘 모른다. 정국이가 알아본 집들 중에 마음에 든 집은 750피트 제곱 크기이며 이 집의 대략적인 가격을 알고 싶어 한다.



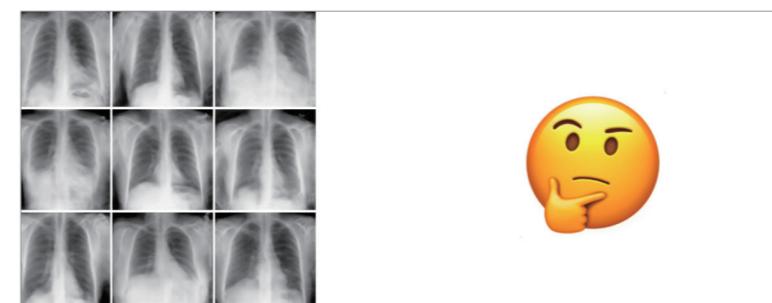
이를 위해 특정 지역의 집 크기에 따른 가격 데이터를 수집했고 데이터 집합을 위 그림과 같이 그래프로 나타냈다. 가로축에는 서로 다른 주택의 크기가 제곱피트 단위로, 그리고 세로축에는 서로 다른 주택의 가격이 천 달러 단위로 표시되어 있다. 그런데 안타깝게도 정국이 집과 같은 크기의 데이터는 없었다. 정국이는 돈이 많지만 적정한 가격에 집을 사고 싶어 한다. 이럴 때 기계학습 알고리즘이 정국이를 도울 수 있을까?



기계학습은 데이터의 경향성을 잘 표현하는 그래프를 찾는 과정이다. 모든 점을 완벽하게 지나지는 않지만 가장 잘 맞는 직선을 그려내는 것이 핵심이다. 기계학습의 결과로 도출된  $h(n)$ 을 사용하면 우리는 데이터에 없던 집의 대략적인 가격을 예측할 수 있을 것이다. 이 데이터를 봤을 땐 15만 달러 정도에 살 수 있을 것으로 보인다.

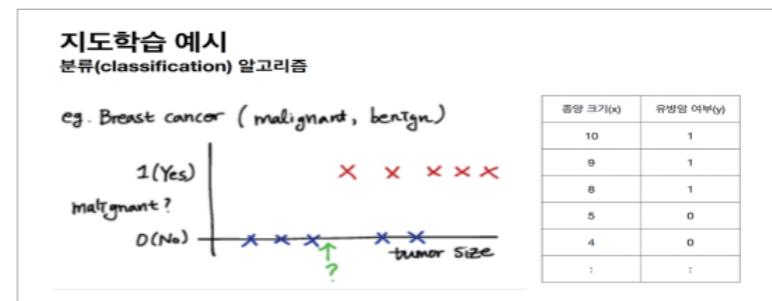
예측 함수(모델) 또는  
가설 함수(모델):  
지도학습의 결과로 도출된  $h(x)$  함수  
를 인공지능 분야에서는 (예측)모델,  
가설로 부른다. 여기서 '모델을 학습  
한다'의 의미는 데이터의 경향성을  
가장 잘 나타내는 함수의 매개변수  
를 찾는 과정을 의미한다.

## ② 지도학습의 동작원리



### 두 번째 예시(분류)

이번엔 환자의 의료 데이터를 보고 유방암이 악성인지 양성인지 예측하고자 한다.



위 그림은 종양크기에 따른 유방암 여부를 데이터로 수집한 후 그래프로 표현한 것이다. 가로축은 종양의 크기이며 세로축은 유방암 여부이다. 종양의 크기가 7인 환자가 있을 때 이 환자는 유방암일 확률이 높을까? 아닐 확률이 높을까? 어렵게도 위 그림에서는 종양의 크기가 7인 데이터는 없다. 그럼 이 문제도 지도학습을 통해 해결해보자. 첫 번째 예시와 마찬가지로 각 데이터의 경향을 잘 나타내는 그래프(함수)를 찾는다. 물론 이 데이터에서는 직선이 아니라 특별한 형태의 함수여야 한다. 길게 늘린 S자 형태면 적당할 것 같다. 그 후 찾아낸 함수에 우리가 알고싶은 종양크기 7을 입력으로 집어넣으면 유방암인지 아닌지에 대한 예측결과가 나오게 된다.



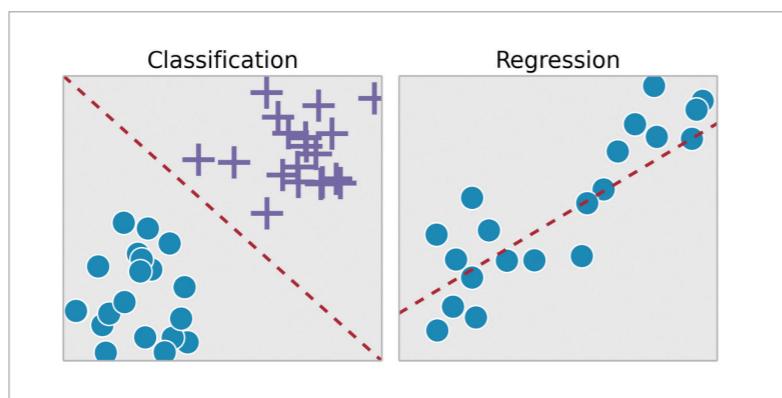
Q. 위 예에서 정국이가 사용한 알고리즘이 지도학습인지 비지도학습인지 쓰고 그 이유를 서술하시오.



Q. 지도학습의 결과로 도출된 함수의 정확도를 높이는 방법은 무엇이 있을지 생각해보시오.

이산 데이터는 개별적이고 분리된 값을 가지며, 일반적으로 셀 수 있는 양을 의미한다. 연속 데이터는 연속적인 값을 가지며, 어떤 범위 또는 스케일 안에서 어떤 값이든 취할 수 있다.

### ③ 회귀와 분류



지도학습의 대표적인 알고리즘으로 회귀와 분류가 있다. 회귀는 연속된 값을 예측하는 문제에서 사용된다. 집값예측 예시에서 집값은 연속적인 숫자 값을 예측하기 때문에 회귀 문제이다. 분류는 이산적인 데이터를 예측하는 것이다. 유방암 예측 예시에서 유방암 여부는 yes or no로 이산적인 값을 갖기 때문에 분류이다.

요약하자면 지도학습의 과정은 정답이 있는 데이터 집합을 준비한 후 데이터를 통해 예측 모델을 학습시킨다. 모델 학습 후 학습하지 않은 새로운 데이터의 결과를 입력하여 값을 추론하고 성능을 평가한다.



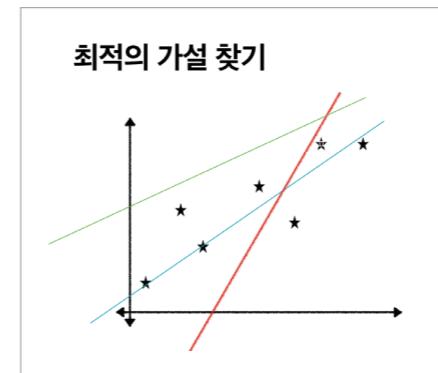
Q. 다음 문제가 회귀와 분류 중 무엇에 해당하는지 쓰시오.

문제	회/분
집값 예측	
손글씨로 쓴 숫자 예측	
인구 수 변화 예측	
고양이 사진 예측	
스팸 메일 예측	

## 3. 비용함수와 경사하강법

### ① 비용함수

이번 시간에는 어떻게 데이터의 경향성을 가장 잘 나타내는 함수(모델)를 찾을 수 있는지 알아보자.



왼쪽 그림을 봤을 때 어떤 색의 함수가 데이터의 경향성을 가장 잘 나타내는 것은 무엇인가? 사람은 눈으로 단번에 답을 찾을 수 있다. 그렇다면 컴퓨터는 어떻게 파란색이 최적인지 알 수 있을까? 컴퓨터는 눈으로 볼 수 없기 때문에 판

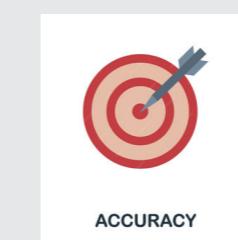
단을 위해서는 기준이 명확해야 한다. 그래서 기계학습에서는 비용(오차)이라는 기준을 두고 최적의 모델을 찾아나간다.

비용함수를 알기 위해서는 비용(오차)이라는 개념이 무엇인지 알아야 한다. 다음 내용부터는 비용과 오차를 같은 뜻으로 사용한다. 지도학습에서 오차란 모델의 예측 값과 실제 값의 차이를 의미하고. 오차 값이 작을수록 예측 값의 정확도가 높아진다고 볼 수 있다. 실제로 모델의 정확도는 다른 여러 가지 요인을 고려하여 산출하지만 지금은 오차만 고려하도록 한다.



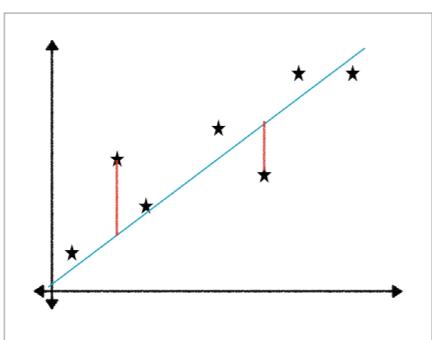
### 모델의 정확도:

모델의 정확도(Accuracy)는 머신러닝에서 모델이 얼마나 잘 예측을 수행하는지를 측정하는 지표 중 하나입니다. 정확도는 주로 분류 문제에서 사용되며, 전체 예측 중 올바른 예측의 비율을 나타냅니다. 간단히 말해서, 모델의 정확도는 다음과 같이 계산할 수 있습니다 예를 들어, 100개의 샘플을 분류하는 모델이 있고, 이 중 90개의 샘플을 올바르게 분류했다면, 모델의 정확도는 90%입니다.



Q. 컴퓨터는 숫자를 좋아합니다. 위 그림에서 3개의 직선의 적합도를 숫자로 어떻게 표현할 수 있을지 생각해 봅시다.

학습률(Learning Rate)의 중요성:  
학습률이 너무 크면 최소값을 지나칠 수 있고, 너무 작으면 학습 속도가 느려지고 지역 최소값에 갇힐 위험이 있다.



비용함수를 구하기 위해 가설  $h(x)$ 를  $y$ 절편과 기울기를  $\theta_0$ 와  $\theta_1$ 이라 할 때 다음과 같은 식으로 표현할 수 있다.

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

다음 변수를 활용하여  $h(x)$ 의 비용함수  $J$ 를 표현하면 다음과 같다.

$m$ : 훈련 데이터 개수

$x$ 's: 학습 데이터의 특징 또는 모델의 입력값

$y$ 's: 학습 데이터의 정답 또는 모델의 출력값

( $x, y$ ): 하나의 훈련 샘플

$(x^{(i)}, y^{(i)})$ :  $i$  번째 훈련 샘플

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

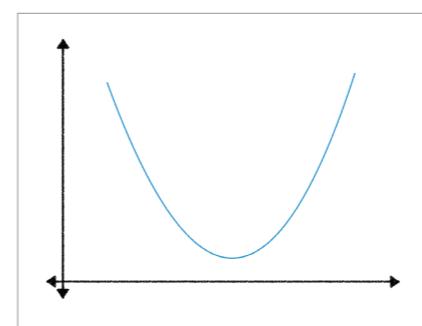


#### Q. 비용함수 계산 프로그램 작성하기.

Hypothesis	
$h_{\theta}(x) = \theta_0 + \theta_1 x$	주어진 가설과 데이터셋을 참고하여 기울기와 바이어스( $y$ 절편) $d$ 의 초기값을 입력받아 비용함수를 계산하여 출력하는 프로그램을 작성하시오.
$x(hours)$	
10	90
9	80
3	50
2	30
$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$	

## ② 경사하강법

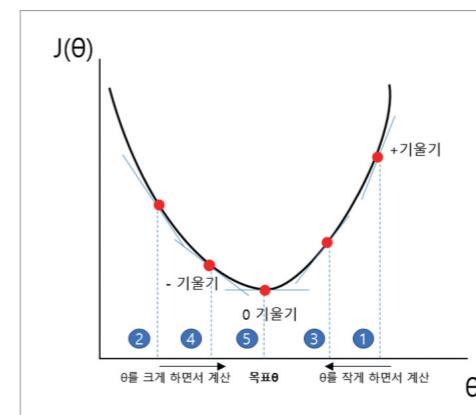
위에서 도출한 비용함수( $J$ )를  $\theta_0$ 에 관하여 표현하면 항상 최고차항의 계수가 양수인 2차 함수인 것을 알 수 있다.



우리는 비용함수를 활용하면 오차 값이 최소가 되는  $\theta$ 를 찾을 수 있다. 그렇다면 어떻게 최저점을 찾는지 알아보자. 사람은 그래프를 보면 최솟값을 바로 찾을 수 있지만 컴퓨터는 값을 기준으로 찾아야 한다. 이 때 사용

되는 알고리즘이 경사하강법이다. 경사하강법의 주요 아이디어는 현재 위치에서 비용 함수의 기울기(즉, 경사)를 계산하고, 그 경사가 감소하는 방향으로 일정 거리만큼 이동하여 함수의 최솟값을 찾아나가는 것이다. 구체적인 방법은 다음과 같다.

$$\theta_{n+1} = \theta_n - (\text{학습률} \times \text{기울기}(\theta_n))$$



1. 초기 값 설정: 매개변수에 대한 초기 값을 임의로 설정. ( $\theta_0$ 를 1번 위치의 값으로 설정)
2. 경사 계산: 현재 매개변수의 위치에서 비용 함수의 경사(미분)을 계산. ( $\theta_0$ 가 1번 위치일 때의 기울기는 양의 값을 갖는다.)
3. 업데이트: 계산된 경사를 사용하여 매개변수를 업데이트 함. 기울기가 양수일 때는 매개변수를 감소시키고 기울기가 음수일 때는 매개변수를 증가시킨다.(1번 위치에서 기울기가 양수이기 때문에  $\theta_0$ 를 감소시킨다.)

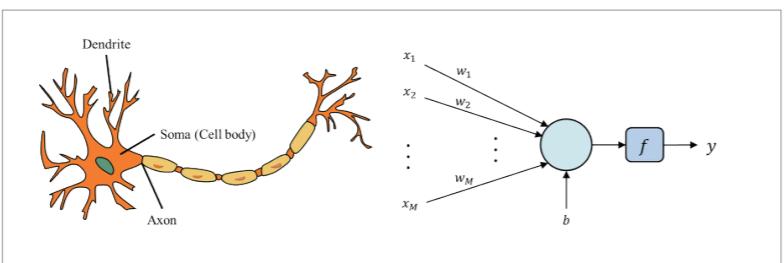
**순전파(Forward Propagation):**  
신경망에서 입력 데이터가 처리되어 출력층까지 이동하는 과정으로, 각 레이어의 가중치와 편향을 이용해 데이터를 처리하고 다음 레이어로 전달합니다. 최종적으로 신경망의 예측값을 산출함.

**역전파(Backpropagation):**  
학습 과정에서 신경망의 오차를 줄이기 위해 사용되는 기법으로, 출력층에서 발생한 오차를 역으로 전파하여 각 레이어의 가중치를 조정함.

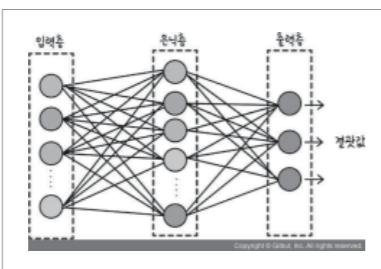
4. 학습률이라는 상수를 사용하여 이동 거리를 조절. 이 때  $\theta_0$ 의 이동 거리는 기울기의 크기와 학습률에 비례한다. (이를 통해  $\theta_0$ 는 2번 위치로 이동하게 된.)
5. 반복: 최솟값에 도달하거나, 더 이상 비용 함수의 값이 크게 변하지 않을 때까지 2와 4의 과정을 반복한다. ( $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$ )

### ③딥러닝

딥러닝은 기계학습 방법 중 하나로 신경망을 사용하여 복잡한 패턴을 찾는다. 신경망은 수많은 노드로 연결되어 있다. 딥러닝의 원리는 위에서 학습한 회귀 알고리즘과 유사하다. 다음 그림을 살펴보자.



노드의 입력은 집값 예측 문제에서 주어지는 다양한 특성들, 예를 들면  $x_1 \sim x_i$ 로 표현되는 집의 크기, 방의 개수 등이다. 여기서  $w_1 \sim w_i$ 는 각 입력 특성에 대한 가중치를 나타낸다. 예를 들어,  $w_1$ 의 가중치 값이 크다면  $x_1$  특성(예: 집의 크기)이 집값 예측에 큰 영향을 준다는 것을 의미 한다. 최적의 예측을 위해, 이 가중치들을 조정하는 과정이 필요하며, 이 때 비용함수를 최소화하는 방향으로 경사하강법을 사용하여  $w_1 \sim w_i$ 의 최적 값을 찾게 된다. 이 과정은 전통적인 회귀 분석에서의 학습 방법과 유사하다.



심층 신경망은 여러 계층으로 노드를 연결시켜 복잡한 문제의 패턴을 학습하는 알고리즘입니다. 심층 신경망은 입력층, 은닉층, 출력층으로 분류되며 순전파와 역전파를 사용하여 최적의 가중치를 학습한다. 심층 신경망의 깊이가 깊어질 때 오차에 대한 정보가 입력층으로 갈 수록 희미해져가는 기울기 손실 문제가 발생하여 학습을 제대로 이루어지지 않는다. 이 문제를 해결하기 위해 활성화함수를 적절히 사용한다.

# 04

## 인공지능 테크놀로지가 뭐지?



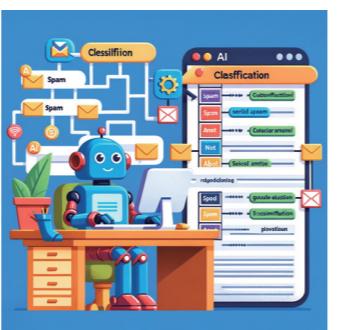
인간이 할 수 있는 일 중 AI가 아직도 어려워하는 것은 무엇일까?

### 1. AI 테크놀로지의 주요 분야



## ① 분류(Classification)

분류는 입력 데이터를 미리 정의된 여러 범주 중 하나로 할당하는 과정이다. 예를 들어, 이메일 필터링에서 스팸 메일과 정상 메일을 구분하는 것이 분류의 사례이며 개인화된 광고 추천 시스템에서 사용자의 웹 행동 데이터나 구매 기록을 기반으로 특정 광고 카테고리를 할당하여 맞춤형 광고를 제공하는 것도 분류작업에 해당한다.



## ② 인식(Recognition)

인식은 데이터에서 특정 패턴이나 객체를 탐지하고 식별하는 과정이며 음성 인식, 얼굴 인식과 같은 기술들이 대표적인 사례이다. 또한 딥페이크 탐지 기술에서 AI는 조작된 이미지나 동영상을 실제와 구분하기 위해 사용되며 스마트홈 기기에서의 화자 인식 기술을 통해 사용자의 목소리를 인식하여 개인화된 서비스를 제공하는 것도 인식의 사례 중 하나이다. 이렇게 인식 기술은 스마트폰, 보안 시스템, 인공지능 비서 등 다양한 분야에서 활용되고 있다.



## ③ 추론(Inference)

추론은 주어진 정보를 바탕으로 새로운 정보나 결론을 도출하는 과정이다. 예를 들어 과거의 기후 데이터를 바탕으로 미래의 날씨 변화를 예측할 수 있다. 또한 의료 분야에서의 새로운 약물을 발견하고자 할 때 AI는 기존의 데이터를 바탕으로 새로운 약물의 효능이나 부작용을 예측하는데 사용되며 사회 네트워크 분석에서는 개인의 소셜 미디어 활동을 기반으로 그들의 관심사나 행동 패턴을 추론하는데에도 사용된다.

## ④ 판단(Decision Making)

판단은 특정 상황에서 최적의 결정을 내리는 과정이며 자율 주행 자동차가 도로 상황을 분석하여 방향을 변경하거나 속도를 조절하는 것이 판단의 한 예이다. 이 외에도 게임 인공지능 로봇 제어, 경영지원시스템 등에서 판단 기능이 활용된다. 최근 금융관련 IT 회사들은 이익을 극대화 하기 위해 신용 점수 산정 시스템에서 AI를 활용한다. 개인의 금융 거래 기록, 지불 이력 등을 분석하여 신용 점수를 결정하거나 대출 승인 결정을 내릴 때 AI의 판단 결과를 사용한다. 또한, 스마트 그리드 시스템에서는 전력 사용 패턴과 예측을 바탕으로 최적의 전력 분배를 결정하는데 AI가 활용되기도 한다.



### Q. 추론과 판단의 차이점을 비교하여 서술하시오.

#### 〈모범 답안〉

"추론"과 "판단"은 실제로 많은 사람들이 혼동하기 쉬운 개념입니다. 두 개념 모두 주어진 정보나 데이터를 기반으로 어떤 결론을 도출하는 과정에 관여하며, 때로는 그 경계가 모호하게 느껴질 수 있습니다. 그러나 명확한 차이점이 있습니다.

#### 1. 추론 (Inference)

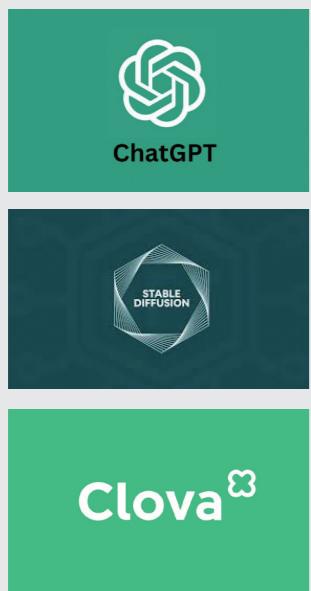
- 정의: 주어진 정보나 데이터를 기반으로 새로운 정보나 결론을 도출하는 과정입니다.
- 예시: 과거의 기후 데이터를 바탕으로 내일의 날씨를 예측하는 것.
- 특징: 추론은 주로 관측되지 않은 상황이나 미래의 사건에 대한 예측을 만들 때 사용됩니다. 이때, 확률론이나 통계를 기반으로 한 계산이 포함될 수 있습니다.

#### 2. 판단 (Decision Making)

- 정의: 특정 상황에서 주어진 여러 옵션 중에서 최적의 선택을 결정하는 과정입니다.
- 예시: 재고 관리 시스템에서 현재의 재고, 판매 예측, 공급망 상황 등을 고려하여 추가 주문을 할지, 기다릴지를 결정하는 것.
- 특징: 판단은 주로 목표나 기준에 따라 최적의 선택을 내리는 과정으로, 여러 대안 중에서 하나나 그 이상을 선택하는 문제 해결 과정에 중점을 둡니다.

#### 차이점 요약

추론은 주어진 정보로부터 새로운 정보나 결론을 도출하는 것에 초점을 맞추며, 판단은 여러 선택지 중에서 최적의 선택을 결정하는 것에 초점을 맞춥니다. 추론이 제공하는 결과는 판단 과정의 한 부분으로 사용될 수 있습니다. 예를 들어, 날씨 예측(추론)을 바탕으로 행사를 진행할지 취소할지(판단) 결정하는 것과 같이 말이죠.



### Q. ChatGPT와 같은 생성형 AI는 어떤 분야에 해당할까?

ChatGPT와 같은 생성형 AI는 여러 AI 분야의 요소를 혼합한 형태로 볼 수 있습니다. 그러나 주요 기능으로는 "인식"과 "추론" 그리고 일정 부분의 "판단" 분야에 해당됩니다.

**인식 (Recognition):** 사용자의 입력 문장을 이해하고 그 의미를 파악하는 과정에서 인식이 중요한 역할을 합니다. 이 과정을 통해 AI는 문장의 주제, 감정, 목적 등을 정확히 파악합니다.

**추론 (Inference):** 입력된 정보와 기존에 학습한 데이터를 기반으로 적절한 응답을 생성하는데 추론 과정이 필요합니다. 예를 들어, 사용자가 특정 주제에 관한 질문을 했을 때, 그 주제에 맞는 적절한 정보나 답변을 도출하는 과정이 추론에 해당됩니다.

**판단 (Decision Making):** 사용자의 질문이나 명령에 대한 여러 가지 가능한 응답 중에서 가장 적절하다고 판단되는 응답을 선택하는 과정에서 판단이 이루어집니다.

다만, 생성형 AI의 경우 "분류"에 직접적으로 해당되지는 않지만, 내부적으로 다양한 분류 작업을 수행하여 사용자의 입력을 이해하고 그에 맞는 응답을 생성하기도 합니다.

따라서, ChatGPT와 같은 생성형 AI는 위에서 설명한 AI 테크놀로지의 주요 분야 중 여러 분야의 특징을 복합적으로 갖추고 있습니다.

## 2. 데이터에서 AI 모델까지

### 공개 데이터셋:

연구 기관, 정부, 대학, 또는 기업들이 특정 목적으로 수집한 데이터를 사용자에게 공개할 때 사용하는 용어이며 이 데이터셋들은 보통 다운로드 가능한 형태로 제공되며, 누구나 접근하여 사용할 수 있다.

### 웹 크롤링:

웹 크롤링은 인터넷상의 웹페이지로부터 자동화된 방법으로 정보를 추출하는 과정이다. 크롤링은 보통 웹사이트의 구조를 분석하여 HTML 문서 내에서 필요한 정보를 찾아내는 방식으로 이루어진다.

### 피쳐 엔지니어링

피쳐 엔지니어링은 주어진 데이터에서 더 유용한 정보나 패턴을 추출하기 위해 데이터의 특성(feature)을 만들거나 수정하는 과정이다. 이는 머신 러닝 모델이 데이터에서 학습하는 데 더 효과적이게 만들어 주며, 모델의 예측 성능을 크게 향상시킬 수 있다.

### ① 데이터로부터 AI 모델을 생성하는 과정



AI 모델을 생성하는 과정은 일반적으로 데이터 수집, 데이터 전처리, 피쳐 엔지니어링, 모델 선택 및 학습, 모델 평가와 같은 6단계의 과정을 거친다. 데이터 수집 단계에서는 AI 모델 생성의 첫 단계로 필요한 정보나 지식을 얻기 위해 원천으로부터 데이터를 모은다. 데이터를 수집하기 위해 공개 데이터셋, 웹 크롤링, 실험 및 시뮬레이션, 데이터 생성 등을 활용한다. 데이터 전처리 단계에서는 수집된 데이터에서 불필요한 정보를 삭제하고 모델 학습에 필요한 형태로 변환한다. 피쳐 엔지니어링 단계에서는 AI 모델의 성능을 높이기 위해 데이터의 주요 특징을 추출하거나 변환한다. 모델 선택 및 학습 단계에서는 적절한 학습 알고리즘을 선택하고 데이터를 사용하여 모델을 학습시킨다. 모델 평가 단계에서는 학습된 모델의 성능을 테스트 데이터셋을 사용하여 평가한다. 마지막 배포 및 최적화 단계에서는 만족스러운 성능을 보인 모델을 실제 환경에 배포하고, 필요에 따라 지속적으로 모니터링 및 최적화한다.



## PACS:

의료 이미지를 전자적으로 저장하고 전송하는 시스템입니다. 병원과 의료 환경에서 주로 사용되며, 의료 영상학(라디올로지)에서 생성되는 다양한 유형의 이미지 데이터를 효율적으로 관리하는데 도움을 줍니다.

## CNN:

Convolutional Neural Network의 약자로 이미지 인식과 처리에 특화된 딥러닝 알고리즘 구조이다. CNN은 이미지의 픽셀 데이터를 입력으로 받아, 다양한 층을 거치면서 이미지의 특징을 추출하고, 이를 바탕으로 이미지 분류, 객체 감지 등의 작업을 수행할 수 있다.

## ② 실제 사례1: 의료 영상을 이용한 병 진단 AI 생성



## ③ 실제 사례2: 온라인 쇼핑몰의 상품 추천 시스템



## 데이터 클린징:

데이터 분석이나 머신러닝과 같은 작업을 수행하기 전에 데이터 세트에서 오류나 불완전한 부분을 수정하거나 제거하는 과정이다. 이는 데이터의 품질을 높여서 보다 정확한 분석이 가능하도록 하며, 결과적으로 더 신뢰할 수 있는 인사이트를 제공한다.

단계	수행 작업
데이터 수집	병원의 PACS에서 다양한 종류의 의료 영상 (MRI, CT, X-ray 등)을 수집한다.
데이터 전처리	영상에서 노이즈를 제거하고, 명도나 대비를 조절하여 통일성을 가진 데이터셋을 생성한다.
피쳐 엔지니어링	영상에서 병의 특징을 잘 표현하는 부분 (예: 종양의 경계, 조직의 밀도 등)을 추출한다.
모델 선택 및 학습	CNN (Convolutional Neural Network)와 같은 딥러닝 모델을 선택하여, 영상과 연결된 진단 정보를 이용해 학습시킨다.
모델 평가	별도로 분리한 테스트 데이터셋을 이용해 모델의 진단 정확도와 다른 성능 지표들을 평가한다.
배포 및 최적화	모델을 병원 시스템에 통합하고, 실제 환자 데이터를 통한 진단에 사용한다. 지속적으로 새로운 데이터를 통해 모델을 업데이트하고 최적화한다.

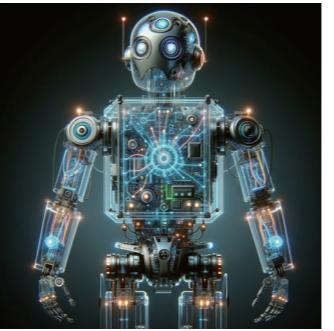
단계	수행 작업
데이터 수집	온라인 쇼핑몰에서 사용자의 구매 이력, 검색 기록, 상품 클릭 기록, 장바구니에 담은 상품 정보 등을 수집한다.
데이터 전처리	사용자의 중복된 데이터 제거, 검색 키워드의 오타 수정, 장바구니에 담았지만 구매하지 않은 상품 정보와 같이 미결제 행동 분석을 통한 데이터 클린징을 수행한다.
피쳐 엔지니어링	사용자의 구매 패턴, 선호하는 상품 카테고리, 가장 최근의 활동, 평균 구매 가격 등의 정보를 기반으로 주요 특징을 추출한다.
모델 선택 및 학습	Matrix Factorization, RNN, 혹은 Transformer 기반의 알고리즘을 선택하여 사용자의 구매 패턴을 기반으로 상품 추천 모델을 학습시킨다.
모델 평가	실제 구매와 모델이 추천한 상품 간의 일치도를 기반으로 모델의 성능을 평가한다.
배포 및 최적화	온라인 쇼핑몰 웹사이트나 앱에 모델을 통합하여 실시간으로 사용자에게 상품 추천을 제공한다. 주기적으로 모델의 성능을 모니터링하고, 새로운 데이터를 통해 업데이트 및 최적화를 진행한다.

### 3. AI 에이전트와 AI 소프트웨어(AISW)

#### ① AI 에이전트와 AI 소프트웨어란?

AI 에이전트는 자기 주변 환경을 살피고, 그 정보를 바탕으로 스스로 결정을 내리고 행동할 수 있는 시스템이다. 예를 들어 스마트폰 인공지능 비서, 인공지능 스피커와 같이 인공지능을 활용하여 물리적 세계와 상호작용하는 것들이 이에 해당한다. AI 에이전트는 이러한 기능을 수행하기 위해 3가지 역할이 필요하다.

- **감지기(Sensor):** 환경에서 다양한 정보를 수집하는 역할(예: 카메라, 마이크)
- **처리기(Processors):** 정보 처리, 학습, 의사결정을 담당하는 역할(예: 인공지능 소프트웨어)
- **행동기(Actuators):** 결정된 행동을 실행하는 역할(예: 모터 제어, 음성 출력)



AI 소프트웨어는 AI에이전트가 잘 동작하게 해주는 프로그램으로 처리기에 해당하며 쉽게 말해, 에이전트의 머리 속 지능이라고 할 수 있다. AI 소프트웨어가 없다면 AI 에이전트는 단순한 컴퓨터 시스템에 불과하다.

#### ② AI 에이전트 동작 과정

AI 에이전트가 사용자에게 서비스를 제공하는 과정은 다음과 같다.

역할	단계	동작
감지기	요청 수신	사용자의 요청을 인터페이스를 통해 수신
처리기	데이터 처리	요청에 따라 필요한 데이터를 수집하고 전처리
	AI 모델 활용	처리된 데이터는 AI 모델에 입력되어 예측, 분류, 추천 등의 작업을 수행
행동기	결과 제공	AI 모델의 결과는 사용자에게 응답 형태로 전달. 이때, 결과는 그래픽, 텍스트, 음성 등 다양한 형태로 표현



#### AI에이전트 vs SW에이전트:

AI 에이전트는 인공지능 기술, 특히 학습과 추론을 통해 복잡한 문제를 해결할 수 있는 능력을 갖춘 소프트웨어 시스템을 말합니다. 이들은 주변 환경에서 데이터를 수집하고, 이를 기반으로 학습하며, 주어진 상황에 대해 최적의 행동이나 결정을 도출합니다. 소프트웨어 에이전트는 사용자의 요구에 반응하거나 자동화된 작업을 수행하는데 중점을 둔 프로그램입니다. 이들은 반드시 고도의 인공지능 기술을 사용하지 않고, 간단한 if-then 규칙이나 사용자 지정 스크립트에 따라 작동할 수도 있습니다. 예를 들어, 이메일 정리를 자동으로 해주는 소프트웨어, 사용자가 정한 조건에 따라 파일을 다운로드하는 에이전트 등이 소프트웨어 에이전트에 속합니다.

### ③ 실제 예시

1) 스마트폰 내 AI 음성 어시스턴트의 날씨 정보 제공



2) 온라인 쇼핑몰의 AI 기반 상품 추천 서비스



역할	단계	동작
처리기	요청 수신	사용자가 스마트폰에 탑재된 AI 음성 어시스턴트에게 "오늘의 날씨는 어때?"라고 물어본다.
	데이터 처리	AI 어시스턴트는 사용자의 위치 정보를 기반으로 가장 가까운 기상 관측소의 데이터를 인터넷에서 수집한다. 또한, 사용자의 위치, 시간대, 계절 등을 고려하여 적절한 날씨 정보를 필터링한다.
	AI 모델 활용	AI는 수집된 데이터를 기반으로 그 날의 기온, 강수 확률, 습도, 바람 속도 등의 정보를 분석한다. 또한, 과거 데이터와 비교하여 특별한 날씨 경고나 조언이 필요한지 판단한다. (예: "오늘은 비가 올 확률이 높으니 우산을 챙기세요!")
행동기	결과 제공	AI 어시스턴트는 분석된 날씨 정보를 사용자에게 응답한다. "오늘 서울의 최고 기온은 28도, 최저 기온은 20도이며, 오후에는 비가 올 확률이 60%입니다. 우산을 챙기시는 것이 좋겠어요!"와 같은 형태로 음성 또는 텍스트로 정보를 제공한다.

역할	단계	동작
처리기	요청 수신	사용자가 온라인 쇼핑몰에 로그인하여 패션 카테고리를 탐색한다.
	데이터 처리	AI는 사용자의 과거 구매 기록, 탐색 기록, 클릭한 상품들, 장바구니에 담긴 상품들 등의 사용자 행동 데이터를 분석한다. 또한, 사용자가 선호하는 브랜드, 색상, 스타일 등의 패턴을 파악한다.
	AI 모델 활용	수집된 사용자 데이터를 기반으로, 현재 쇼핑몰 내 어떤 상품들이 사용자의 취향에 가장 맞을지 예측한다. AI는 상품의 이미지, 설명, 리뷰 등의 다양한 정보를 분석하여 추천 상품 리스트를 생성한다.
행동기	결과 제공	사용자가 쇼핑몰의 메인 페이지나 상품 페이지를 탐색하면, "당신을 위한 추천 상품"이라는 섹션에서 AI가 선정한 상품들을 표시한다. 이를 통해 사용자는 자신의 취향에 맞는 상품을 더 쉽게 발견할 수 있게 되며, 쇼핑몰은 매출 증대에 기여하게 된다.



Q. 음악 스트리밍 서비스의 AI 기반 개인화된 플레이리스트 추천 시스템에서 에이전트의 동작과정을 단계별로 쓰시오.



역할	단계	동작
감지기	요청 수신	
처리기	데이터 처리	
	AI 모델 활용	
행동기	결과 제공	

## II. AI와 데이터, 윤리



1. 데이터가 왜 중요하지?
2. 데이터를 어떻게 다루지?
3. 윤리문제의 종류와 발생원인은?
4. 윤리 문제 해결을 위한 교육, 그리고 법과 사회제도

## 01

# 데이터가 왜 중요하지?



## 1. 데이터의 이해

### ① 데이터의 정의

데이터는 분석 혹은 참조하기 위해서 모은 값이며 유의미한 관측값들의 모임이다. 데이터는 숫자, 영자 혹은 주기(period), 정부(+, -) 부호의 특수 문자에 의해 구성되며 디지털의 기본 단위로서 0과 1의 이진법으로 표기된다.



출처	정의
옥스퍼드 영어사전	참조 또는 분석을 하기 위해서 모아놓은 사실 그리고 통계값
영문 위키피디아	한 사람 또는 여러 사람 아니면 한 개체 또는 여러 개체에 대해서 모은 숫자 혹은 질적인 값의 모임

### ② 빅데이터의 개념

빅데이터란 생성주기가 짧고 방대하며 아날로그 방식으로 수집, 저장, 분석, 시각화가 어려운 정형 데이터 혹은 비정형 데이터로 글자, 이미지, 영상 등 모든 요소가 함축된 대규모 데이터이다.



### ③ 빅데이터의 중요성

4차 산업혁명은 정보 통신 기술의 융합으로 이루어지는 차세대 사회 변화라고 할 수 있다. 4차 산업혁명 중 단연 가장 중요한 키워드는 ‘빅데이터’이다. 4차 산업혁명에서 언급하는 모든 기술은 데이터 기반으로 이루어진다. 인공지능, 로봇공학, 사물인터넷, 무인 운송 모두 빅데이터가 꼭 필요하다.

인공지능의 핵심은 지금까지 사람이 해왔던 의사결정 및 판단을 인공지능 모델로 대체해 더 정확하고 빠르게 자동으로 하는 것이다. 인공지능 모델은 주로 머신러닝과 딥러닝 기술을 활용해서 만들어지는데, 딥러닝의 가장 핵심이 되는 것이 빅데이터다. 빅데이터 안의 데이터가 다양할수록 그 안에 세상의 패턴 또는 정보가 많이 담기기 때문에 훨씬 더 정확한 모델을 갖출 가능성을 지닌다.

① **정보 가치 최대화**: 빅데이터는 다양한 정보를 포함하고 있어 기업이나 기관이 가지고 있는 정보의 가치를 최대화할 수 있다. 이를 통해 고객의 행동 패턴 파악이나 시장 동향 예측 등을 통해 비즈니스 전략을 개선할 수 있다.

② **의사 결정 지원**: 빅데이터 분석은 조직이나 기업의 의사 결정을 지원합니다. 데이터 분석을 통해 문제 해결이나 새로운 기회를 찾을 수 있으며, 이를 통해 조직의 성과를 향상시킬 수 있다.

③ **혁신과 경쟁력 강화**: 빅데이터는 혁신과 경쟁력을 강화하는 데에도 중요한 역할을 한다. 데이터 분석을 통해 새로운 제품, 서비스, 비즈니스 모델을 개발하거나 기존 프로세스를 최적화할 수 있다.

④ **예측과 예방 가능성**: 빅데이터 분석은 과거 데이터를 기반으로 미래를 예측하고 예방할 수 있는 가능성을 제공한다. 예측 모델을 구축하여 잠재적인 문제를 조기에 감지하고 예방 조치를 취할 수 있으며, 이를 통해 리스크를 최소화하고 효율성을 높일 수 있다.

데이터(data)는 ‘datum’의 복수형의 컴퓨터 용어로 정보를 작성하기 위해 필요한 자료를 뜻한다

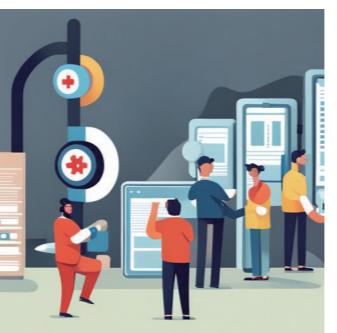
데이터  
관찰이나 측정을 통해 수집한 값 또는 사실

데이터 속성  
데이터를 특성에 맞게 분류하는 기준

4차 산업혁명  
기계학습과 인공지능의 발달로 인한 산업의 변화를 가리키는 말  
예: 3D프린터, 로봇, 스마트폰, 우주 공학

산업혁명 정리  
1차: 증기기관 기반의 기계화 혁명  
2차: 전기 에너지 기반의 대량생산  
3차: 인터넷 기반의 지식정보 혁명  
4차: 정보기술과 빅데이터의 결합

⑤ 산업 및 사회 발전: 빅데이터는 산업 및 사회 발전에 기여한다. 의료 분야에서는 질병 예방이나 정확한 진단, 개인 맞춤형 치료 등을 실현할 수 있고, 교통, 에너지, 환경 분야에서는 지능형 시스템을 구축하여 지속 가능한 발전을 이룰 수 있다.



빅데이터의 중요성은 데이터의 양과 다양성이 계속 증가하고 있기 때문에 데이터를 효과적으로 수집, 관리, 분석하는 능력이 기업이나 기관의 경쟁력을 결정하는 핵심 요소로 간주된다.

#### ④ 빅데이터의 활용성

분야	데이터수집	데이터활용	주요기술	AI프로그램(위튼:wrtn)을 사용하여 생성한 그림
로봇	• 범위, 위치, 시각, 촉각 센서 등에서의 빅데이터 수집 • 머신러닝, 딥러닝을 활용한 데이터 분석	• 수집된 데이터를 활용하여 목적과 책임을 결정 • 머신러닝, 딥러닝을 활용한 계획 수립	빅데이터 분석, 머신러닝, 딥러닝	
사물인터넷	• 온습도, 빛, 풍향, 초음파, GPS 등의 센서에서의 데이터 수집 • 네트워크 기술을 통한 데이터 송수신	• 수집된 데이터를 분석하여 환경 정보 획득 • 데이터를 활용한 지능적인 서비스 제공	센싱 기술, 네트워크 기술, 데이터 분석	
드론	• 다양한 위치에서의 데이터 수집 • 다양한 센서를 통한 데이터 수집	• 다양한 데이터를 효과적으로 수집 • 데이터 분석을 통한 다양한 정보 획득	센싱 기술, 데이터 수집 및 분석	
의료 및 헬스케어	• 환자 정보, 검사 결과 등의 데이터 수집 • 센서를 통한 생체 신호 수집	• 질환 진단 및 치료 보조 소프트웨어 개발 • 음성 인식을 통한 서비스 제공	인공지능 기술, 음성 인식, 데이터 분석	



#### 빅데이터 활용 사례를 적어보자.

예) 서울시에서는 통화량 데이터를 분석하여 심야에 운영하는 올빼미 버스 노선을 만들었다.



활동지

#### 기계학습에 필요한 데이터 속성을 찾기

건강 관리 에이전트에게 나의 일별 운동량을 학습시키려고 한다.  
하루 동안 접한 데이터 중에서 기계학습에 필요한 데이터 속성을 찾아 정리해보자.

##### 1. 오늘 하루 동안 내가 접한 데이터를 모두 기록해보자.

예) 날씨 애플리케이션: 오늘의 기온, 습도  
동영상 스트리밍 애플리케이션: 동영상 시청

##### 2. 나의 하루 운동량을 확인하려고 한다. 1에서 작성한 내용을 바탕으로 필요한 데이터와 필요하지 않은 데이터를 분류해보자.

필요한 데이터	필요하지 않은 데이터
예) 계단 오르내리기 10분	예) 동영상 스트리밍 웹 사이트 순위

##### 3. 나의 하루 운동량을 표로 만들어보자.

① 표로 만들 때 고려해야 할 데이터 속성을 적어 보자.

예) 운동, 시간(분), 소모 칼로리 등

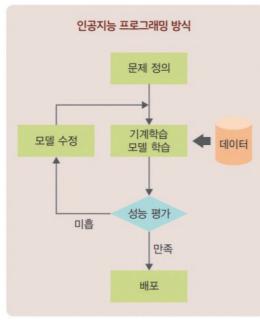
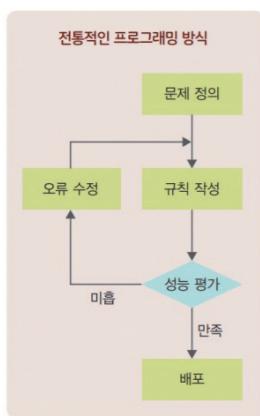
② ①에서 적은 데이터 속성을 기준으로 하루 운동량 표를 완성해보자.

시각	운동	시간(분)	소모 칼로리
예) 9:30AM	예) 계단 오르내리기	예) 10	예) 61

**지식의 종류**  
선언적 지식: 세설, 개념 원리 등에 관한 지식  
절차적 지식: 어떤 이링 절차나 방법, 규칙에 관한 지식

**추론의 방법**  
연역적 추론: 현재 가지고 있는 지식을 조합해 새로운 지식을 발견하는 것  
귀납적 추론: 몇 가지 사례로부터 지식을 발견하는 것

## 2. 데이터에 기반한 인공지능 원리



### ① 세상의 모든 문제를 인공지능을 활용할 수 있을까?

세상의 수많은 문제 중 컴퓨터나 인공지능을 활용하여 해결할 수 없는 문제에는 어떤 것이 있을까? 가족 간의 사랑처럼 마음에 대한 문제, 새로운 분야의 문제를 발견하거나 정의하는 문제 또는 서로 다른 영역을 서로 연결하는 문제 등은 컴퓨터나 인공지능으로 해결하기 어려울 것이다.



인공지능을 활용하여 해결할 수 있는 문제와 그렇지 않은 문제를 구분  
지어 생각해보자.

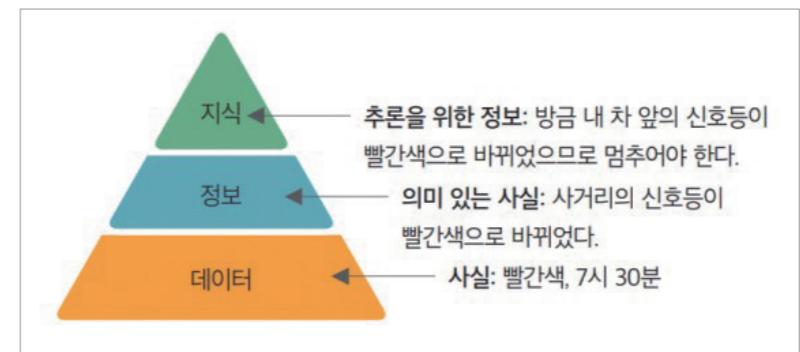
예)

① 해결 가능한 문제: 예) 드론 및 딥러닝을 이용한 자율 주행 자동차용 3D HD Map

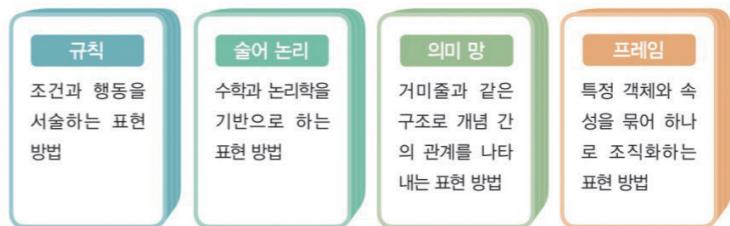
② 그렇지 않은 문제: 예) 사칙연산과 같은 계산 문제

### ② 지식 표현과 추론

지식이란 어떠한 사항에 대한 명확한 인식이나 이해 또는 알고 있는 내용을 말한다. 보통 외부 세계의 여러 데이터를 가공하여 어떤 의미를 갖도록 한 것을 정보라 하고, 이 정보를 취합하고 분석하여 문제 해결에 사용할 수 있도록 이해한 것을 지식이라고 한다. 추론은 이미 알고 있는 지식을 가지고 새로운 사실을 이끌어 내는 지적 능력이다. 인공지능 연구 초창기부터 추론을 가장 주요한 지능의 하나로 보고 컴퓨터가 자동으로 추론하도록 만드는 방법이 활발하게 연구 되었다.



지식 기반의 인공지능 개발에 관심이 집중되면서 컴퓨터를 이용해 처리하고 저장할 수 있는 형태로 지식을 표현하는 여러 가지 방법이 고안되었다.



[그림] \_지식의 표현 방법

특정 분야에서 전문가의 의사 결정 능력을 흡내 내는 프로그램을 전문가 시스템이라고 하는데, 핵심 구성 요소에는 사용자 인터페이스, 지식 베이스 추론 엔진이 있다. 사용자 인터페이스는 의사가 문진하듯이 주로 대화 형식으로 만들어지는데, 전문가 시스템의 질문에 사용자가 답을 입력하면 입력된 내용으로 전문가 시스템이 추론한 결과를 설명 준다. 지식 베이스는 전문가의 지식을 저장하며, 지식 베이스가 더 잘 만들어질수록 전문가와 더 유사한 의사 결정을 내릴 수 있을 것이다. 마지막으로 추론 엔진은 지식 베이스를 이용해 새로운 사실을 추론한다.

챗GPT

G	Generative 생성형 인공지능 모델
P	Pre-trained 사전 훈련
T	Transformer 핵심적인 신경망 모델

### ③ 데이터 기반한 지능의 형성

챗GPT는 궁금한 것을 텍스트로 입력하면 인공지능이 딥러닝을 통해 사람이 적은 듯한 글들을 만들어내는 언어 모델이다. 챗GPT가 광범위한 질문에 대답할 수 있게 하는 가장 중요한 요소는 데이터다. 챗GPT 시대에서 데이터는 예전과 비교할 수 없을 정도의 거대한 양의 데이터가 필요하다. 따라서 데이터 수집, 정제, 분석, 관리 등의 데이터 관련 기술이 매우 중요해지고 데이터를 지속적으로 관리하고 분석함으로써 성능을 개선할 수 있다.



오픈 AI는 인터넷, 뉴스 기사, 문서, 채팅 기록 등 다양한 데이터 소스를 활용하여 대규모의 데이터를 수집한다. 이렇게 수집된 데이터들은 정제 과정을 거쳐 챗GPT의 학습에 사용된다. 학습 과정을 통해 챗GPT는 입력된 데이터를 분석하고 패턴을 파악하여 새로운 데이터를 생성할 수 있는 능력을 향상한다.

챗 GPT를 포함한 생성형 인공지능은 텍스트, 이미지, 음성 등 다양한 데이터를 학습하는데 데이터의 양과 질이 인공지능의 성능을 좌우한다. 따라서 데이터 수집, 정제, 분석, 관리 등의 데이터 관련 기술이 매우 중요해진다. 데이터 없이는 인공지능이 존재할 수 없다.

02

## 데이터를 어떻게 다루지?



### 1. 데이터 리터러시

#### ① 데이터 리터러시

최근에는 다양한 데이터에서 가치를 뽑아내 해석하고 시각화하고 잘 전달하는 능력까지도 포함하여 데이터를 문맥에 맞게 일고 사용하고 소통하는 능력은 물론이고 데이터에 담겨 있는 의미를 파악하고 목적에 맞게 활용하는 능력을 의미한다. 일반적으로 리터러시는 ‘무엇을 할 수 있는 능력’으로도 해석되어 기술 리터러시, 디지털 리터러시, 미디어 리터러시, 비주얼 리터러시 등이 광범위하게 사용된다.



데이터 리터러시	데이터를 읽고 쓸 줄 아는 능력 'data'와 'literacy'의 합성어
----------	--

## ② 정형 데이터와 비정형 데이터

정형 데이터	구조화되어 있고 열과 행으로 이루어진 테이블 형태로 저장	대부분 예측 가능한 패턴을 가짐. 대부분의 전통적인 데이터베이스 시스템에서 사용되며 SQL 쿼리를 사용하여 집계, 조인 등 다양한 연산 수행 가능
비정형 데이터	구조가 없거나 자유로운 형태로 저장되며 주로 문서, 이미지, 음성 등 다양한 형태의 컨텐츠를 포함	복잡하고 동적인 내용을 담고 있음. 기계 학습 및 자연어처리, 이미지·음성 인식과 같은 고급 분석 기술을 활용하여 의미 있는 정보를 추출하고 분석한다.

따라서 정형 데이터와 비정형 데이터 모두 중요하지만 각각 다른 방식으로 처리하고 분석해야 한다. 최근에는 정형 데이터와 비정형 데이터를 함께 활용하는 것도 중요한 추세이다.



### 정형 데이터와 비정형 데이터 분류하기

수지는 친구와 함께 영화관에서 영화를 보려고 한다. 어떤 영화를 볼지 고민하다가 어떤 영화가 가장 인기가 많았는지 조사해 보기로 하였다.

1. 어떤 영화가 인기가 많은지 조사할 때 필요한 데이터는 무엇인지 적어보자.

2. 위에서 작성한 데이터를 정형 데이터와 비정형 데이터로 분류해 보자.

정형 데이터	비정형 데이터
예) 일일 관람객 수	예) 영화 감상 후 올린 티켓 사진

## ③ 데이터와 인공지능의 관계

인공지능은 문제를 해결하기 위해 수많은 데이터를 활용하며, 이 데이터를 활용해 만들어 낸 정보는 또 다른 데이터가 된다. 이 과정을 반복하는 동안 인공지능은 세상의 수많은 정보를 학습하여 미래를 예측하고 인간의 의사 결정에 영향을 미치기도 한다.



지도학습에서는 정답이 존재하는 데이터를 필요로 하며, 일반적으로 정형 데이터를 활용한다. 데이터 속성이 존재하는 정형 데이터는 인공지능을 학습하는 데 활용되며 데이터 속성을 통해서 사용자가 원하는 답을 얻을 수 있다.

비지도 학습은 정답을 얻기보다는 방대한 데이터 사이의 패턴을 발견하여 비슷한 유형의 데이터끼리 묶는 것으로, 일반적으로 비정형 데이터를 활용한다. 비정형 데이터는 데이터 속성이 존재하지 않기 때문에 인공지능 프로그램이 수많은 비정형 데이터를 비교·분석하여 특징을 찾아 기준을 세우고 데이터를 정형화한다. 즉 비정형 데이터의 패턴이 비슷한 것끼리 묶어 데이터로 활용하는 것이다.

## ④ 데이터 수집방법

① 설문 조사: 설문지나 질문지를 활용하여 직접 응답자로부터 데이터를 수집하는 방법입니다. 온라인이나 오프라인으로 설문 조사를 실시할 수 있으며, 응답자의 의견과 피드백을 수집할 수 있다.



② 관찰: 관찰은 특정 대상의 행동이나 현상을 직접 관찰하여 데이터를 수집하는 방법입니다. 예를 들어, 실험실에서 동물의 행동을 관찰하거나 상점에서 고객의 구매 패턴을 관찰하는 것이 있다.

③ 인터뷰: 인터뷰는 직접 대상과 대화하여 정보를 수집하는 방법입니다. 구조화된 인터뷰와 비구조화된 인터뷰가 있으며, 응답자의 의견과 경험을 자세히 파악할 수 있다.

④ 공공 데이터 활용: 공공기관이나 공공기관에서 제공하는 데이터를

### 지도학습

문제와 정답이 주어진 상태에서 학습시키는 방법

### 비지도학습

답을 주지 않고 데이터의 특징을 스스로 발견하는 방법

### 공공 데이터

공공 데이터는 정부에서 제공하는 데이터이다.



공공 데이터는 공공 데이터 포털 ([www.data.go.kr](http://www.data.go.kr)), 통계데이터센터 ([data.kosrar.go.kr](http://data.kosrar.go.kr)), 기상자료개발포털 ([data.kma.go.kr](http://data.kma.go.kr)) 등에서 제공하고 있다.

데이터를 활용한 심야 버스 노선 최적화: 올빼미 버스



활용하여 분석할 수 있습니다. 정부 통계, 공공기관의 오픈 데이터 포털 등에서 다양한 데이터를 활용할 수 있다.

⑤ 웹 스크래핑: 웹 스크래핑은 웹페이지의 정보를 추출하여 데이터를 수집하는 방법입니다. 웹 스크래핑 도구를 사용하여 특정 웹사이트의 내용이나 데이터를 수집할 수 있다.

⑥ 센서 데이터 수집: 센서를 통해 환경이나 사물의 데이터를 실시간으로 수집할 수 있습니다. 예를 들어, 기상 관측 장비를 통해 기온, 습도 등의 데이터를 수집하거나, 스마트 홈에서 가전제품의 사용 데이터를 수집할 수 있다.

⑦ 로그 데이터 수집: 로그 데이터는 시스템이나 애플리케이션에서 발생하는 이벤트나 행위에 대한 기록을 말합니다. 서버 로그, 애플리케이션 로그 등을 분석하여 사용자 동작 패턴 등의 정보를 수집 할 수 있다.

데이터 수집 방법은 수집하려는 데이터의 특성과 목적에 따라 선택되어야 한다. 데이터 수집 시에는 개인정보 보호와 관련된 법률과 윤리적인 측면을 고려하여 적절한 방법을 선택하고 수집 절차를 따라야 한다.

## ⑤ 데이터 시각화

데이터 시각화는 데이터 속성을 도형, 선, 색 등의 시각적 요소를 활용하여 표현하는 것을 말한다. 이는 데이터에서 규칙이나 패턴을 찾아 분석한 것을 이해하기 쉬운 상태로 재구성하는 것을 의미하며, 오래전부터 인간이 정보를 효과적으로 전달하기 위해 사용한 의사소통 방법이다. 시각화의 종류에는 분포 시각화, 시간 시각화, 비교 시각화, 관계 시각화, 공간 시각화 등이 있다.

유형	설명	예시
분포 시각화	데이터의 분포와 특성을 시각적으로 나타냄	히스토그램, 박스 플롯, 산점도
시간 시각화	데이터의 분포와 특성을 시각적으로 나타냄	선 그래프, 시계열 그래프, 캘린더 히트맵
비교 시각화	다른 항목 또는 그룹 간의 비교를 시각화하여 상대적인 차이를 보여줌	막대 그래프, 원 그래프, 히트맵
관계 시각화	데이터 간의 상관관계나 연관성을 시각화하여 관계를 파악함	네트워크 그래프, 트리맵, 산포 행렬
공간 시각화	지리적 정보를 시각화하여 공간적 패턴과 분포를 이해함	지도 시각화, 도표 지도, 공간 클러스터링



'캠핑'이라는 단어의 성별 및 연령에 따른 검색량 비교 분포 시각화를 보고 내용을 분석해보자.



- 남성과 여성 중 어느 쪽이 캠핑을 더 선호하며, 이는 어떤 차트를 보고 분석할 수 있는지 적어보자.
- 위 차트를 보고 어떤 연령층이 캠핑을 선호하는지 적어보자.



공공데이터를 활용하여 데이터를 시각화하고 분석해보자.

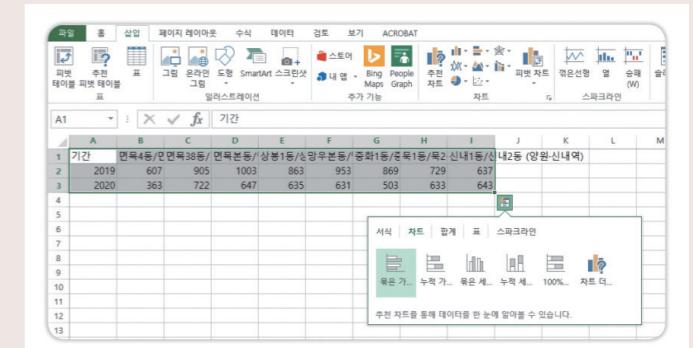
1. 공공데이터 포털 웹 사이트에 접속하여 필요한 데이터를 다운로드한다.

- ① 공공데이터포털 웹 사이트([www.data.go.kr](http://www.data.go.kr))에 접속하여 키워드를 검색하거나 [데이터 찾기] 탭 등을 활용하여 원하는 데이터를 검색한다.
- ② '안심 스카우트' 데이터를 검색하여 다운로드 한다.



2. 다운로드한 데이터 파일을 열어 시각화해 보자.

- ① 스프레드시트 형태인 데이터 파일(csv 또는 xls 확장자)를 다운로드하여 전용 프로그램에서 연다.
- ② 시간의 변화, 여려 요소 간의 비교, 장소의 변화 등과 같은 조건에 따라 어떤 방식으로 시각화할지 정한다.
- ③ 필요한 데이터를 모두 드래그한 다음 [삽입] 탭의 [차트] 기능을 활용하여 차트를 만든다.



## 분류 문제의 해결 과정



## ⑥ 분류 문제의 해결 과정

분류란 일정한 기준을 가지고 주어진 대상을 나누는 활동을 말한다. 기계학습은 주어진 데이터 속성의 학습을 통해 최적의 결과를 산출하기 위해 문제해결 모델을 만든다. 이 과정에서 데이터의 속성을 규칙이나 패턴에 따라 분류하여 인공지능에 사용할 모델을 만든다. 그 다음에 새로운 데이터를 입력하면 사전에 학습된 모델을 통해 주어진 데이터의 규칙이나 패턴을 분석하여 적절히 분류할 수 있다. 이처럼 학습한 규칙을 통해 판단하는 방법을 분류 모델이라고 한다. 분류 모델의 성능은 기계학습에서의 성능과 밀접한 관계가 있다. 다양한 데이터를 학습하여 만든 분류 모델을 이용하면 새로운 입력 데이터를 분류할 수 있다.

문제 발견 및 정의하기	새로운 문제를 발견하고 문제를 명확하게 정의하는 단계이다. 우리가 해결할 수 있는 문제인지 아닌지, 인공지능으로 해결할 수 있는 문제인지 아닌지를 판단한다.
데이터 수집하기	인공지능으로 해결 가능한 문제라면 우선 필요한 데이터를 수집한다. 데이터를 직접 만들거나 측정하여 수집할 수 있고, 이미 준비된 데이터 세트를 사용할 수도 있다.
데이터 탐색하기	수집된 데이터의 통계 정보를 확인하거나 다양한 시각화를 통해 데이터의 속성을 파악한다. 이를 통해 학습에 사용할 속성과 제외할 속성을 결정한다.
데이터 전처리하기	데이터 탐색 과정에서 확인한 누락된 값이나 이상한 값을 처리하고, 학습에 적합한 형태로 데이터를 가공한다. 자연어 데이터를 형태소나 품사 단위로 처리할 수도 있고, 이미지 데이터의 경우 부족한 데이터를 늘릴 수도 있다. 이러한 과정을 '전처리'라고 하며, 일반적으로 전체 문제 해결 과정 중 많은 시간이 소요된다.
모델 선정 및 학습하기	일반적으로 분류 문제에 적합한 모델과 적합하지 않은 모델이 있지만, 모든 문제에 완전히 맞는 모델은 존재하지 않는다. 따라서 다양한 모델을 후보로 생각하고 문제와 데이터의 특성에 맞는 모델을 선정하여 학습한다.
평가 및 모델 개선하기	학습 결과를 평가한 후, 평가 결과를 바탕으로 다른 모델로 문제를 해결하거나 같은 모델로 설정을 달리하여 성능을 개선한다.
배포하기	학습된 모델을 검증했다면 실제 문제를 해결할 수 있도록 모델을 다른 사람들과 공유한다.

03

## 윤리 문제의 종류와 발생 원인은?



Q. 우리가 아는 인공지능 기술은 어떤 것들이 있을까요?

Q. 위에서 언급한 인공지능 기술이 갖는 긍정적인 기능과 부정적인 기능을 함께 이야기해 봅시다.



순기능	인공지능 기술	역기능

## 1. 인공지능 윤리

## ① 윤리

윤리란 '사람으로서 마땅히 행하거나 지켜야 할 도리'이다. 우리 사회의 여러 기능 중 법으로 강제할 수 없는 영역에서 사람이 믿고 있는 가치에 따라 스스로 판단하여 행동하도록 만드는 일종의 암묵적인 규율이다. 그러므로 윤리는 인간이 사회를 구성하고 살아가는 데 있어서 이치라고 볼 수 있다.

## 인공지능의 특성

- 빅데이터
- 인식
- 학습
- 추론
- 예측
- 문제 해결

## ② 인공지능 윤리란?

인공지능 윤리는 인공지능을 개발·운영·사용함에 있어 개발자와 소비자에게 요구되는 윤리 의식을 의미한다. 인공지능 윤리는 사람이 속한 사회의 윤리 규범과 큰 틀에서는 같은 개념을 갖지만, 인공지능에 의해 발생하는 윤리적 문제들은 인간 사회에서 일어나는 문제들과는 다른 방식으로 일어나므로 인공지능의 기술적 원리와 윤리적 기준을 동시에 이해하는 것이 필요하다. 인공지능은 기술적 특성으로 인해 인간의 개입이 없어도 스스로 학습할 수 있으며, 그 결과를 반드시 인간이 예측할 수 없는 경우도 있다. 앞으로 더 많은 인공지능 기술이 우리의 삶에 직접적인 영향을 줄 수 있으므로 인공지능의 기술이 어떻게 사용되는지에 대해 주의를 기울일 필요가 있다.



더 알아보기

### 윤리 vs 법



“우리 나라는 전 국민의 완벽한 안전을 위하여 오늘부터 24시 개인을 지키는 인공지능 기술을 도입합니다. 개인을 따라다니면서 관찰하는 인공지능 드론을 도입하여 각 개인을 위험으로부터 보호함과 동시에 위법한 행위를 예방할 것입니다.

이 인공지능 드론은 개인이 편안한 속면을 취하고 있는지, 건강상의 문제는 없는지 계속 관찰하게 됩니다. 야간에 어둡고 인적이 드문 곳을 걷더라도 주변의 위험 요소를 빠르게 감지하여 영상을 녹화하고 경찰이 충돌하게 함으로써 사람들을 보호할 수 있습니다.

반면, 개인이 범법 행위를 했을 때에는 누가 어디서 어떤 위법을 했는지 널리 알려 다른 사람들도 범법 행위를 하지 않도록 경각심을 불러일으키도록 할 것입니다.”

**인공지능을 윤리가 아닌 법으로만 통제한다면 세상은 어떻게 될까요?**

## 2. 인공지능 윤리 문제

### ① 인공지능 윤리 문제

인공지능이 우리 사회에 미치는 영향이 커지고 있다. 인간이 추론하기 어려운 영역을 단기간 내에 빠르게 추론하여 합리적인 의사결정을 돋고, 작게는 인공지능 기반 맞춤형 추천을 통해 개인의 취향을 파악하고, 네트워크를 연결할 수 있는 전자기기를 똑똑하게 제어하는 등 많은 긍정적 효과를 제공하고 있다. 그러나 자율 주행 자동차의 사고, 혐오나 차별 발언을 하는 챗봇, 인공지능으로 만든 가짜 뉴스 등 부정적 영향도 나타나고 있다.

이러한 인공지능의 양면성으로 인해 인공지능을 활용하는 사회에서 다양한 윤리적 쟁점이 발생한다. 이러한 윤리적 쟁점은 대부분 사회 구성원(개발자, 사용자, 운영·관리자)이 인공지능을 개발하고 활용하는 과정에서 발생한다. 따라서 우리는 사회적 책임과 인공지능 사회에 걸맞은 윤리 의식을 갖고 공정성을 추구하는 방향으로 인공지능을 개발하고, 사용하며, 운영 및 관리해야 한다.

#### 인공지능 추천 서비스 예시

- 넷플릭스, 유튜브 등의 스트리밍 영상 서비스
- 멜론, 지니, 플로, Spotify 등 음악 서비스
- amzon, 네이버, 롯데온 등 쇼핑 서비스

#### AI 전자기기 제어 서비스 예시

- 음성 제어 서비스, 스마트 홈 시스템, 스마트 시티 시스템, 자율 주행 자동차 등

#### 챗봇 ‘이루다’의 사례



이루다는 스캐터랩사의 일상대화형 인공지능 챗봇으로 윤리적 법적 문제를 제기해 출시 3주를 넘기지 못하고 서비스가 종료되었다. 편향성, 개인 정보 침해, 무분별한 AI 기술 남용, 성적 논란 등의 윤리적 문제를 남겼다.

#### 영국 대학시험 알고리즘 차별 논쟁 사례

영국 대입 시험에서 학습자 거주지 역에 따른 알고리즘 차별로 교육 불평등 강화의 논란을 남겼다.

#### 미국 교사평가 알고리즘 투명성 논란 사례

공립학교 교사 고용에 있어 비밀 알고리즘이 평가를 실시하여 교사 연맹이 소송을 제기한 사건이다.

구성원	윤리적 쟁점	실천 방안
개발자	<ul style="list-style-type: none"> <li>편견이나 오류가 있는 데이터를 인공지능의 훈련에 사용함으로써 인공지능의 실수나 편향성을 유발할 수 있다.</li> <li>인공지능을 설계할 때 개발자의 선입견이나 편견이 작용할 수 있다.</li> </ul>	<ul style="list-style-type: none"> <li>데이터나 관련 기술 설계 등에 편향성이 포함되지 않도록 주의한다.</li> <li>인공지능 시스템이 결정을 내린 과정을 이해할 수 있도록 설명이 가능한 형태로 개발한다.</li> </ul>
사용자	<ul style="list-style-type: none"> <li>인공지능을 활용하는 과정에서 개인 정보가 동의 없이 수집될 수 있다.</li> </ul>	<ul style="list-style-type: none"> <li>자신의 개인 정보를 삭제할 수 있는 법적 권리가 있음을 알고, 필요할 경우 삭제를 요구한다.</li> </ul>
운영·관리자	<ul style="list-style-type: none"> <li>인간이 제어하기 어려울 정도로 인공지능 시스템이 진화할 수 있다.</li> <li>인공지능 기술의 비싼 사용 비용으로 특정 계층만 사용할 경우 사회적 불평등을 심화할 수 있다.</li> </ul>	<ul style="list-style-type: none"> <li>인공지능 시스템의 작동, 운영을 인간이 관리 및 감독 할 수 있도록 한다.</li> <li>많은 사람이 인공지능 기술을 사용할 수 있도록 보장하고, 인공지능 기술의 혜택이 공정하게 분배되게 한다.</li> </ul>

[표] 인공지능 사회 구성원별 윤리적 쟁점과 사회적 책임, 공정성 추구를 위한 실천 방안 예시

## [윤리 원칙 타임라인]

2017

- 딥마인드 윤리&사회 원칙 (딥마인드)
- 아실로마 인공지능 원칙 (삶의 미래 연구소)

2018

- 인공지능 원칙(MS, 구글, IBM)
- AI 알고리즘 윤리 헌장(카카오)
- 지능정보사회 윤리 가이드라인 (정보문화포럼)

2019

- 인공지능에 대한 이사회 권고안 (OECD)
- 신뢰할 수 있는 인공지능을 위한 윤리 가이드라인(EU)
- 매일의 인공지능 윤리(IBM)
- 인공지능 윤리& 안전에 대한 이해 (앨런 튜링 연구소)
- 인공지능 국가 전략(법정부)
- 이용자 중심의 지능 정보사회를 위한 원칙(방송통신위원회)
- 로봇 윤리 기본 원칙 (산업통상자원부)
- 인공지능 윤리 협장 (한국인공지능 윤리협회)

2020

- 인공지능 윤리를 위한 로마의 요구 (로마 교황청)
- 인공지능 공정성 체크 리스트 (MS)
- 신뢰할 수 있는 인공지능을 위한 자체 평가 목록(EU)
- 사람이 중심이 되는 인공지능 윤리 기준(법정부)
- 자율 주행차 윤리 가이드라인 (국토교통부)

2021

- 인공지능 윤리에 대한 권리 (UNESCO)
- 신뢰할 수 있는 인공지능 실현 전략(안)(방송통신위원회)
- 인권과 안전, 민주주의가 보장되는 인공지능 정책을 요구하는 시민 사회 선언(국내 120개 시민 단체)
- 네이버 AI 윤리 준칙(네이버)
- 인공지능 추구 가치(SKT)

인공지능의 윤리적 쟁점을 정확하게 규정하기는 어려우나, 세계 경제 포럼 등 국제 단체나 관련 기관에서 일반적으로 제시하는 윤리적 쟁점은 다음과 같다.

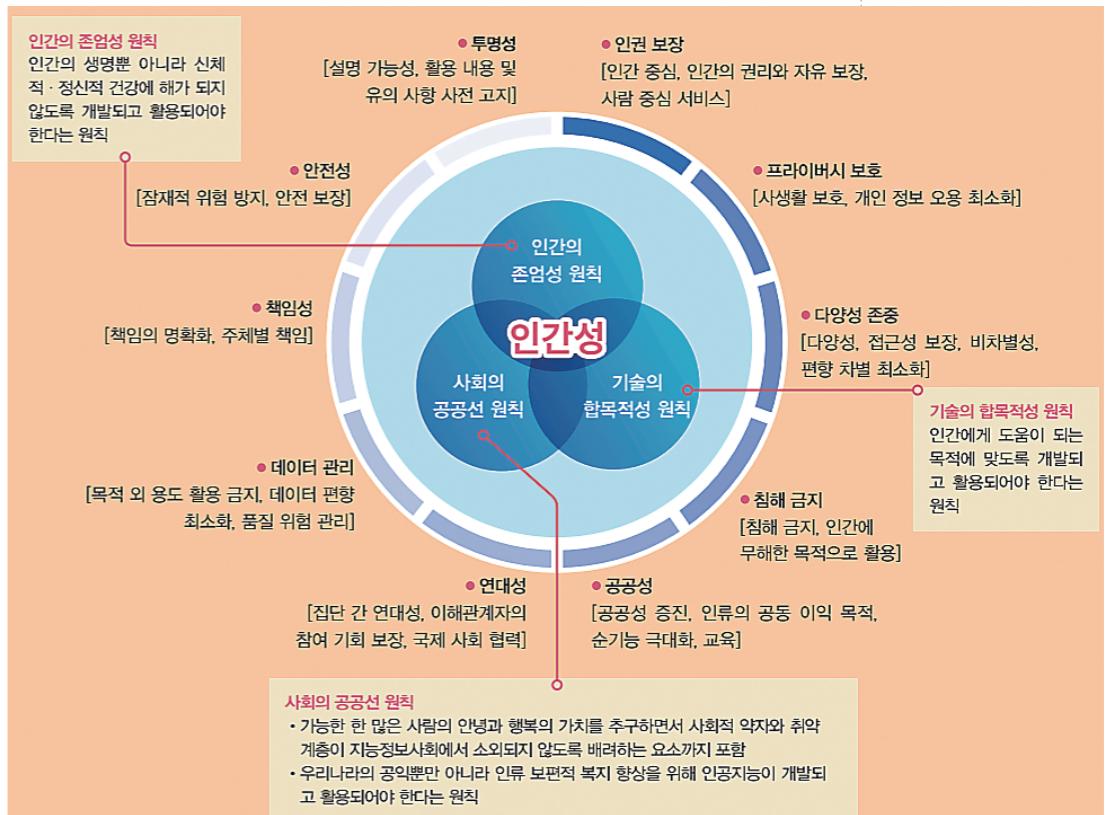


구글 포토 서비스가 흑인 여성을 고릴라로 인식하여 인종차별 문제가 발생하였다.  
(출처: <https://blog.naver.com/musmamedia/222099174823>)

## ② 바람직한 인공지능 사회를 만들기 위한 노력

인공지능 윤리에 대한 기존의 논의는 일자리 문제나 사용자의 오남용에 집중되어 있었다. 그러나 현재는 개발부터 데이터를 수집하고 알고리즘을 공정하게 설계하고, 사용자의 권리를 보호하기 위한 방법을 준비하는 등 전 과정에서 고려해야 할 윤리적 문제를 파악하여 구체적인 제도를 마련하는 방향으로 나아가고 있다. 인공지능이 인간에 의해 만들어지고, 활용하는 주체도 인간이기 때문에 인공지능 개발 단계부터 지침을 만들고 개발자의 표준 책무를 제시하여 인공지능이 어떤 특성을 가지고 행동해야 하는가를 사전에 고려함으로써 윤리적 쟁점을 최소화하는 것이 중요하다.

따라서 바람직한 인공지능 사회를 위해서 인공지능 개발자, 사용자, 운영·관리자 등 인공지능 사회 구성원 모두가 책임 있는 윤리 의식을 갖추고 공정성을 추구해야 한다. 그러기 위해서는 인공지능의 기술적 특성과 사회적 영향을 고려한 윤리 기준이 필요하다. 우리나라에서도 이에 따라 사람이 중심이 되는 「인공지능 윤리기준」을 만들어 3대 원칙과 10대 핵심 요건을 통해 인공지능 시대에 대응하려고 준비하고 있다.



인공지능 윤리 기준의 기본 원칙과 핵심 요건

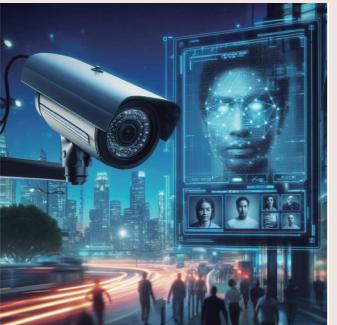
(출처: 「인공지능 윤리」, 과학기술정보통신부 및 정보통신정책연구원.)

- 2022
- 인공지능 권리장전 AI Bill of Rights(미 백악관)
  - 인공지능 개발과 활용에 관한 인권 가이드 라인(국가인권위원회)
  - 교육분야 인공지능 윤리 원칙 (교육부)
  - AI 윤리 원칙(LG)
  - AI 첫봇 윤리 준칙(스캐터랩)



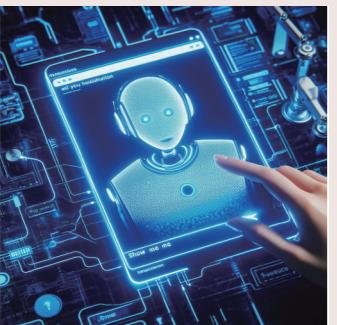
다음 자료를 읽고 인공지능 윤리에 대해 토론해 봅시다.

### 자료1



안면 인식 기술은 휴대전화의 잠금장치를 편하게 해제하거나 집을 잃은 치매 노인을 찾는데 도움을 제공하기도 합니다. 그런데, 인공지능이 잘못된 판단을 하여 문제가 생기기도 합니다. 미국 경찰이 안면 인식 기술을 이용하여 범죄자를 빠르게 체포하려는 과정에서 무고한 시민이 범죄자로 오인되어 체포되고 구금되는 사건들이 발생했습니다. 각 사건에서 범인으로 지목된 사람은 모두 흑인으로, CCTV 카메라에 잡힌 그들의 모습이 절도범의 모습과 유사하다는 인공지능의 판단에 따라 적게는 30시간 많게는 10일간 교도소에 구금되었습니다.

### 자료2



A 사의 인공지능 비서가 한 가족의 대화를 녹음하여 사용자의 지인에게 이를 전송하는 일이 벌어졌습니다. 이 사용자는 지인으로부터 전화를 받고 사적인 대화가 전송되었음을 알게 되었는데, 알고 보니 인공지능 비서가 사용자의 대화를 잘못 알아듣고 “OO에게 메시지 전송”이라는 단어로 인식하여 대화를 녹음한 후 전송한 것이었습니다.

Q1. 사람은 자신의 실수에 대해 책임을 집니다. 예를 들어, 자동차를 운전하다가 실수로 사고가 났을 때 사과를 하고 치료비를 지원합니다. 만약 인공지능 때문에 사고가 발생한다면 이 책임은 누구에게 있는 것일까요?

Q2. 인공지능도 윤리가 필요한 것일까요? 아니면 인공지능을 만든 사람이 윤리적 태도를 갖추어야 하는 것일까요?

# 04

## 윤리 문제 해결을 위한 교육, 그리고 법과 사회 제도



AI 이론



유럽연합(EU)이 세계 최초로 인공지능(AI) 규제법 초안을 마련했다. 글로벌 차원의 AI규제안을 마련하려는 발걸음이 빨라지고 있다.

유럽의회 산하 소비자보호위원회는 2023년 5월 11일(현지시간) EU 전역에서 챗GPT, 미드저너 등 AI체계를 규제하기 위한 법안 관련 입장을 채택했다.

Q: 왜 글로벌 차원에서 AI를 규제하려고 하는 것일까요?

### 1. 인공지능 윤리 문제와 교육

국내외에서 인공지능과 관련하여 많은 문제가 발생하고 있다. 우리나라에서는 교육적 차원에서 문제의 심각성을 인지하고 최근 교육부에서 ‘사람의 성장을 지원하는 인공지능’이라는 대원칙 아래 ‘학습자의 주도성 강화’, ‘교수자의 전문성 존중’, ‘기술의 합목적성 제고’라는 3대 기본 원칙과 사람, 공동체, 기술을 결합한 교육 분야 인공지능 윤리 세부 원칙을 9가지로 제시하였다. 또한 교육부에서는 AI윤리이슈를 발굴하고, AI윤리교육을 강화시킬 것이며, 안정성을 판단할 수 있는 도구 개발과 현장수





요맞춤형 AI개발을 지원하겠다고 밝혔다.

교육 분야에서의 인공지능 활용 가능성에 대해서도 많은 연구가 이루어지고 있다. 학습 관리부터 평가, 교사 지원을 비롯하여 평생 교육에 이르기까지 많은 가능성이 제시되고 있다. 활용 가능성이 높아지는 만큼 인공지능 윤리 문제가 발생하지 않도록 개발, 사용, 운영에 있어서 적절한 윤리 가이드라인이 계속 연구되어야 할 것이다.

#### 참고 교육분야 인공지능 윤리원칙

대원칙		사람의 성장을 지원하는 인공지능		
기본 원칙		학습자의 주도성 강화	교수자의 전문성 존중	기술의 핵심적 성 제고
세부 원칙	사람 + 공동체 + 기술	교육분야 인공지능은		
				1. 인간성장의 잠재기능성을 이끌어낸다. 2. 모든 학습자의 주도성과 다양성을 보장한다. 3. 교육당사자 간의 관계를 공고히 유지한다. 4. 교육의 기회균등 실현을 통해 공정성을 보장한다. 5. 교육공동체의 연대와 협력을 강화한다. 6. 사회 공공성 증진에 기여한다. 7. 모든 교육당사자의 안전을 보장한다. 8. 데이터 처리의 투명성을 보장하고 설명 가능해야 한다. 9. 데이터를 핵심적으로 활용하고 프라이버시를 보호한다.

인공지능은 이제 특정한 분야나 집단에서만 활용하는 것이 아니기 때문에 모두가 인공지능 기술 발전의 혜택에서 소외되지 않고, 인공지능이 발달하면서 발생하는 문제들을 해결하기 위해서는 전국민적 차원에서의 교육을 강화하여 관심과 참여를 유도할 필요가 있다.

인공지능(AI)은 사람의 생각과 말을 학습해 답을 예측하는 알고리즘이다. 학습 데이터가 올바르지 않으면 잘못된 결과를 도출하게 되고 이는 사회문제로 이어진다. AI개발자와 사용자 모두 교육을 통해 사회에 대한 책임 의식을 가진다면 인공지능이 더 나은 세상을 만드는 기술로 발전할 수 있을 것이다.

#### 참고 교육에서 인공지능 활용 가능 분야

분류	기능	예시
학습 관리	· 출석, 과제 제출 관리 · 학업 중단 위기 학생 관리	· 교육용 챗봇
학습/평가	· 학생 맞춤형 학습/평가 지원 · 양질의 교육에의 접근 보장	· 지능형 교육 시스템(ITS) · 대화 기반 교육 시스템(DBTS)
교사 지원	· 자동 과제 관리 시스템 · 자동 평가 시스템 · 자동 토론 시스템	· 인공지능 교사, 인공지능 교실
평생 교육	· 인공지능을 통한 평생교육 시스템 구축	· 인공지능 평생학습 안내자 · 인공지능 평생학습 포트폴리오



#### 함께 해봐요

Q. 학생들은 인공지능을 이미 경험하고 있고, 향후 개발하거나 운영·관리를 하게 될 수 있는 사회 구성원입니다. 자신의 교과와 관련하여 인공지능 윤리 문제를 예방을 위한 수업을 구상해 봅시다.

구분	교과 내용
교과명	

## 2. 인공지능 윤리 문제와 법

기능에 따른 AI 분류

### 1. 반응 기계(Reactive Machines)

미래의 행동을 위해 기억이나 과거 경험을 경험하지 않으며 현재 시나리오에만 집중하고 가능한 최선의 조치에 따라 반응한다.

ex) 구글의 알파고

### 2. 한정된 기억(Limited Memory)

과거 경험이나 일부 데이터를 짧은 기간동안 저장하고 한정된 기간 동안만 저장된 데이터를 사용한다.

ex) 자율 주행 자동차

### 3. 마음 이론(Theory of Mind)

인간의 감정, 사람, 신념을 이해하고 인간처럼 사회적 상호 작용을 하는 AI로 아직 개발되지는 않았다.

### 4. 자아 인식(Self-Awareness)

자아 인식은 인간만이 경험하는 지능 발달 단계로 자기 정체성 또는 개성을 경험하는 것이다. 자아 인식은 아직 현실의 존재하지 않는 가상의 개념이다.



인공지능의 기술 발전은 빠르게 이루어지고 있지만, 이 기술을 어느 수준까지 허용할 것인지에 관한 사회적 논의는 아직 초기 단계에 머물러 있다. 법과 제도는 강제력이 있으나, 만들거나 개정하는 데에 시간이 많이 소요되기 때문에 빠르게 발전하는 기술을 규제하기에는 한계가 있다.

또한 업계에서는 기술 발전에 저해가 될 수 있다는 주장이 제기되기도 한다. 따라서 인공지능 윤리에 대한 사회 구성원들의 합의가 먼저 이루어지고 나서, 윤리 기준에 크게 위반된다거나 큰 부작용을 불러오는 기술의 경우에 법으로 규제하는 것이 바람직 할 것이다.

인공지능의 윤리적·법적 문제에 대한 대응은 미국, 유럽 연합(EU) 등을 비롯한 국제사회에서 지속적으로 논의되었다. 우리나라에서도 2019년 과학기술정보통신부를 비롯한 관계부처의 참여로 「인공지능 국가전략」이 마련되었고, 2020년에는 「인공지능(AI) 윤리 기준」이 제시되었다. 이외에도 한국정보화진흥원의 「지능정보사회 윤리 가이드라인」, 국가인권위원회의 「인공지능 개발과 활용에 관한 인권 가이드라인」이 있다.

이렇듯 정부, 국회, 여러 위원회, 시민단체 등에서 인공지능 법제화 및 행정지침 마련을 위한 시도들이 있었으나 접근 방식이 급진적이었고, 주체마다 다른 양상을 보이는 상황이다. 따라서 인공지능 관련 법제 및 행정 지침 마련은 인공지능 서비스의 발전 속도를 살펴보고 공론화 과정을 충분히 거친 뒤 마련되어야 할 것이다. 다만 인공지능의 위험성은 이전과는 다른 복잡성을 갖고 있기 때문에 위험을 방관하지 말아야 하며 개별 서비스에서의 경험을 토대로 일관된 통제 기준을 촉적해나가야 할 것이다.

일반적 입법 규율은 다양한 인공지능 기반 서비스에 적용하기 어려운 측면이 있기 때문에 인공지능의 '기능'에 따라 구분해 해당 기능에 탑재된 제품이나 서비스에 두루 적용되도록 필요가 있다. 공통적으로 적용되는 필수 불가결한 원칙들을 일반법에서 정하고, 나머지는 개별 인공지능서비스에서의 위험 양상의 특성을 검토해 수정, 반영할 수 있도록 민간협력 또는 자율규제에 맡길 수 있을 것이다.

마지막으로 업계의 자율적 참여로 개별 인공지능서비스에 맞춤형 정책이 필요하다. 여러 나라에서 기술 위험에 대응해서 인류와의 공존을 위한 인공지능 윤리를 마련했지만, 그 이면에는 기술을 선점하고자 하

는 이해관계가 얹혀 있다. 이 때문에 인공지능 윤리지침은 개별 인공지능 응용서비스에 적용하기에는 추상적이고 내용상 부족한 부분이 있다. 또한 윤리 가이드라인과 윤리현장은 개발자와 공급자, 이용자가 준수할 지침들을 제시하고 있지만, 지능정보사회를 예측한 수준에서 상정한 개념들을 행동 지침으로 마련하고 있어서 앞으로 출시되는 인공지능서비스에서는 실효성이 낮을 가능성이 있다. 따라서 개발·운영사들이 윤리 가이드라인 및 규범을 만드는 데에 참여함으로써 전문가집단 및 규제기구가 함께 공동의 가이드라인을 마련하는 것이 적절할 것이다.

### 참고 AI 관련 해외 입법 동향

국가	입법 동향	내용
미국	알고리즘책임법안 입법 추진(2022년 3월)	자동화된 의사결정 시스템에 대한 책임성 확보 위해 자체적인 알고리즘 영향평가를 실시·제출하도록 의무 부과
	통신정보청(NTIA) 인공지능 책임성에 대한 공개 여론 수렴(2023년 4월)	인공지능 책임성, 투명성, 책임주체 관련 여론 조사 합리적 규제 수준 도출을 위해 인공지능 책임성 규제 수준에 대한 대중과 기업의 인식 조사
	인공지능 규제 백서 발간(2023년 3월)	혁신 도모 및 규제 불확실성 감소 등 성장과 번영 추진 공공 신뢰 제고 법 규제 최소화 및 규제-혁신추진 간 균형 확보 분야별 유연하고 합리적 규제 추진
영국	AI 및 데이터 보호 가이드 일부 개정(2023년 3월)	책무성과 거버넌스 투명성 및 적법성 등 일부 내용 추가 인공지능 규제백서에 맞춰 변하는 환경에서 AI의 미래지향적 활용 위한 개정
	인공지능법안(AI Act, 2021년 4월)	AI 위험 수준 4가지로 구분해 수준에 따른 규제 부과
	AI Act 관련 EU 이사회 수정의견(2022년 11월)	국가안보, 연구목적 및 비전문적 영역에 대한 적용 제외 고위험 AI의 적용범위 등 분류체계 일부 수정 고위험 인공지능 규제의 요건 준수의 기술적 실현 가능성 향상 및 수법자 부담 경감 혁신 지원 대책 강화
	AI Act, 유럽의회 통과(2023년 6월)	생성형 AI 규제, 공공장소에서 원격 생체 인식 금지 등

### AI 위험 수준에 따른 규제(유럽)

#### 1. 수용불가 (Unacceptabel Risk)

인간 행동을 조작해 사용자 자유의지를 방해하는 AI 시스템 사용 금지  
ex) 음성비서를 사용하는 장난감이 미성년자에게 위험한 행동 독려

#### 2. 고위험 (High Risk)

위험을 최소화할 데이터세트를 갖춰야 하며, 이용자에게도 AI 시스템의 위험성에 대해 충실히 공지해야 하며 위험을 최소화하기 위해 인간이 적절한 관리 감독할 것  
ex) 중요한 인프라스트럭처(교통 등), 신용점수 평가, 수술 보조 프로그램, 법 집행 등

#### 3. 제한된 위험 (Limited Risk)

첫봇 같은 AI 시스템을 이용할 경우에 이용자들이 기계와 소통하고 있다는 사실을 인식하도록 할 것

#### 4. 최소 위험 (Minimal Risk)

AI 기술을 자유롭게 활용 가능  
ex) 비디오 게임, 스팸 차단기 등



1. 인공지능 기술을 사용할 때, 발생할 수 있는 문제 사례 중 1가지를 선택하여 탐구해 봅시다.

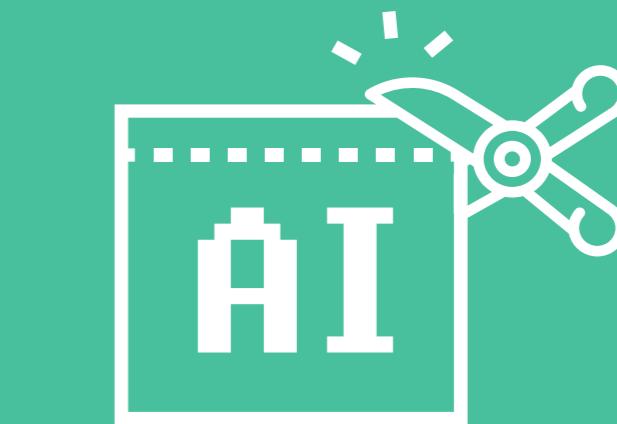
#### EU 보고서에서 제시한 문제 사례

- |   |  |
|---|--|
| 1 | 신입 사원을 모집하는 데 사용되는 알고리즘은 일반적으로 여성보다 남성을 선호하는 것으로 나타났다. |
| 2 | 온라인 챗봇은 몇 시간 만에 인종 차별주의자가 됐다.                          |
| 3 | 기계 번역은 젠더 편향을 보였다.                                     |
| 4 | 안면인식 시스템은 백인 남성을 잘 감지하지만, 흑인 여성의 경우에는 그렇지 않았다.         |

구분	내용
선택한 문제 사례	
문제 사례의 개요	
침해된 기본 권리 요소	

2. 탐구한 기본권 문제에 대한 해결책을 기술적·제도적·윤리적 차원으로 제시하고 토론해 봅시다

구분	기술적 차원	제도적 차원	윤리적 차원



## AI 앱 제작실습



# I. AI로 데이터 시각화 분석



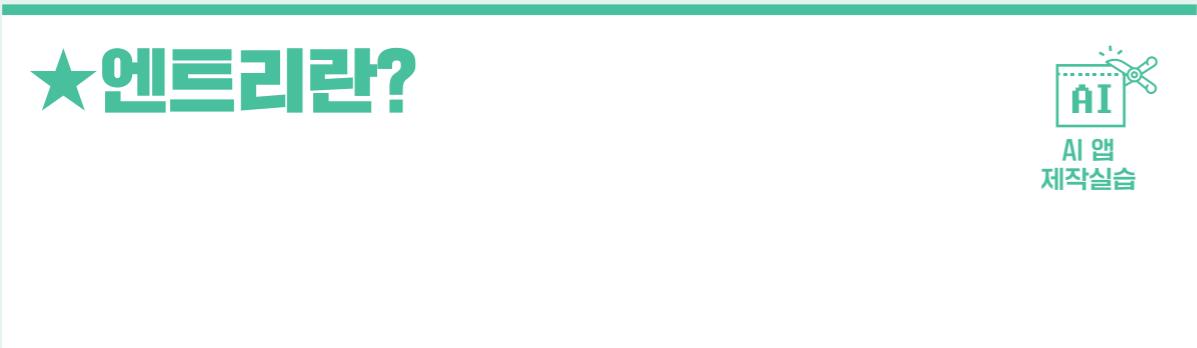
이번 단원에서는 엔트리를 이용하여 데이터를 전처리, 분석하는 방법을 배우게 됩니다. 엔트리를 이용한다면 블록 코드를 활용하여 쉽게 데이터를 처리, 분석할 수 있습니다. 이번 내용을 통해 정성적으로만 확인했던 기존의 실험 과정에 정량적인 데이터 분석을 더한다면 보이지 않았던 새로운 관계를 파악할 수 있을 것입니다.

## ★ ‘엔트리’란?

1. AI로 데이터를 시각화 및 분석해 보자(엔트리)

## ★ ‘오렌지3’란?

2. AI로 데이터를 시각화 및 분석해 보자(오렌지3)



엔트리 사이트 주소  
<https://playentry.org/>

## 1. 엔트리 소개

엔트리를 활용하여 생각하고, 만들고, 공유하는 활동을 할 수 있다.

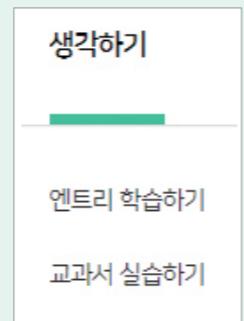


## 2. 생각하기, 만들기, 공유하기

### ● 생각하기

소프트웨어를 통해 다양한 경험을 할 수 있고, 학생들의 역량을 찾아 줄 수 있다. ‘생각하기’ 활동을 통해 다양한 인공지능 기술을 활용할 수 있다. ‘생각하기’는 ‘엔트리 학습하기’와 ‘교과서 실습하기’가 있다.

‘엔트리 학습하기’에서는 엔트리에서 제공하는 간단한 미션을 통해 코딩의 기초를 학습할 수 있다. ‘교과서 실습하기’에서는 각 출판사에 실린 교과 내용을 코딩을 통해 실습해 볼 수 있다.



### ● 작품 만들기

블록 코딩을 통해 상상만 했었던 작품, 게임, 도구 등을 직접 만들어 볼 수 있다. ‘작품 만들기’에는 ‘작품 만들기’, ‘교과형 만들기’, ‘스터디 만들기’가 있다.

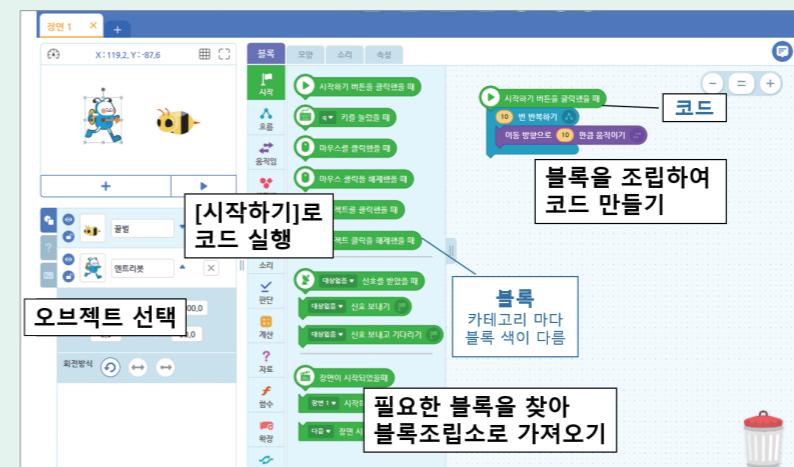


‘작품 만들기’ 화면을 살펴보면 다음과 같다.



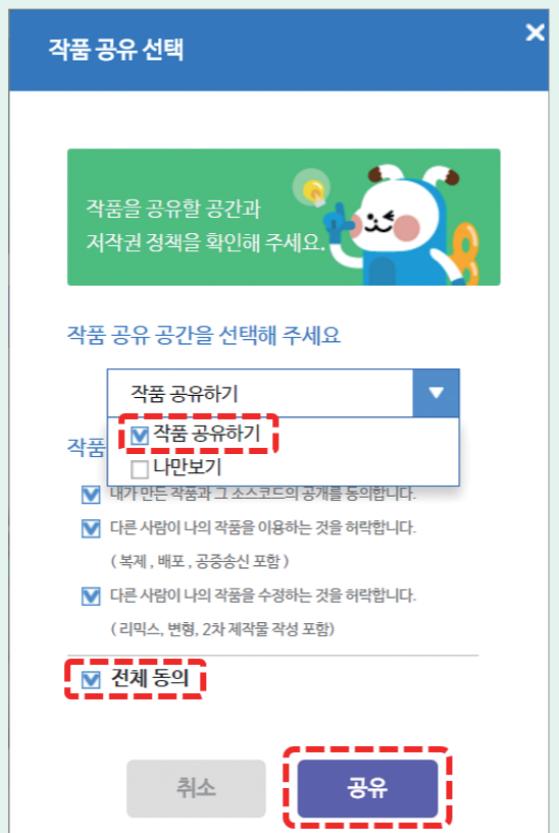
화면은 ‘실행화면’, ‘오브젝트 목록’, ‘블록 꾸러미’, ‘블록 조립소’로 구성되어 있으며, ‘블록 꾸러미’에서 블록을 가져와 ‘블록 조립소’에서 코드를 만드는 방식이다.

만든 코드를 이용하여 오브젝트에게 명령을 내린다.



### ● 작품 공유하기

마이페이지에서 공유 버튼을 눌러 만든 작품이 '공유하기'에 업로드 된다.



01

## AI로 데이터를 시각화 및 분석해 보자(엔트리)

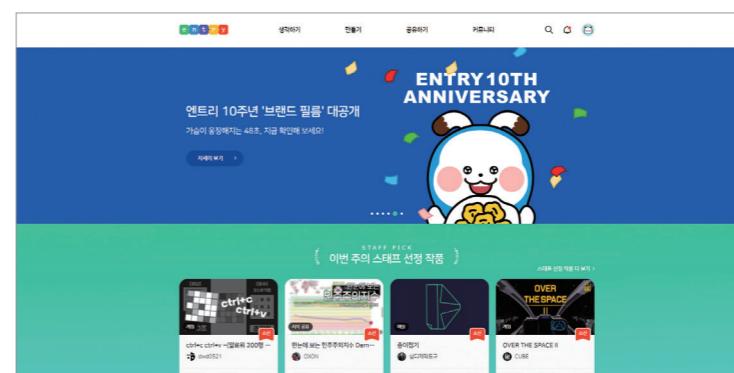


엔트리 사이트 주소  
<https://playentry.org/>

### 1. 데이터 업로드 및 확인하기

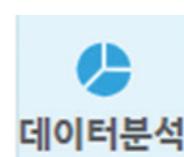
#### ① 데이터 업로드하기

- '엔트리' 사이트에 로그인을 하고, 상단의 '만들기' - '작품 만들기' 탭을 선택한다.



(연습용) Train Data  
<https://zrr.kr/yXas>

- '엔트리' 메인 화면 중 '데이터 분석'을 클릭하고, '데이터 불러오기'를 클릭 한다. (좌측의 연습용 데이터를 다운로드 받는다.)



(연습용) Test Data  
<https://zrr.kr/bPLT>

- '테이블 불러오기' 화면 좌측에 '테이블 추가하기' 버튼을 클릭한다.



'엔트리' 메인 화면 중 '데이터 분석'을 클릭하고, '데이터 불러오기'를 클릭 한다.(좌측의 연습용 데이터를 다운로드 받는다.)

데이터는 CSV 확장자로 다운로드 한다.

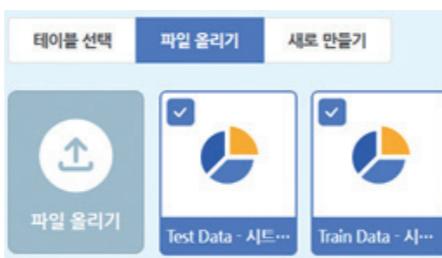
#### CSV란?

Creating Shared Value의 약자로 쉼표를 기준으로 항목을 구분하여 저장한 데이터

- 상단에 '테이블 선택', '파일 올리기', '새로 만들기' 버튼이 있다. 학습 상황에 맞는 탭을 선택한다.(본 내용에서는 '파일 올리기'를 선택한다.)



- 다운로드 받은 'Train Data'와 'Test Data'를 업로드하고, '추가하기' 버튼을 클릭한다.



## ② 데이터 확인하기

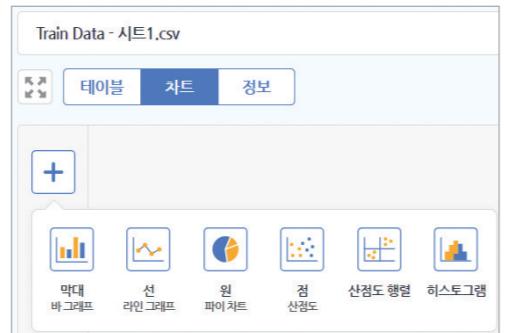
- 업로드한 테이블을 선택하면 우측 화면에서 데이터를 확인할 수 있다.

	A	B	C	D	E
1	날짜	요일	공부한 시간	시험 점수	공부를 시작한 시간
2	2023-03-02	목	2	55	8:00:00
3	2023-03-03	금	3	69	8:03:00
4	2022-12-28	수	5	74	8:10:00
5	2022-11-28	월	7	83	8:05:00
6	2022-11-01	화	8	95	8:13:00

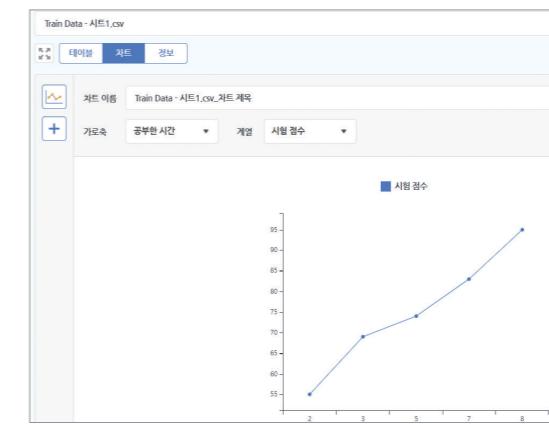
연습용 데이터는 '선 그래프'로 경향성을 확인한다.

## ③ 데이터를 차트로 확인하기

- '차트' 버튼을 클릭하고, '+' 버튼을 클릭하면 그림 종류를 선택할 수 있다.



- 가로축과 세로축을 설정하면 그림을 확인할 수 있다.



## 데이터 전처리하기

### ① 테이블 정리하기

공부한 시간에 따라 쉬는 시간을 제공하려고 한다. 쉬는 시간에 대한 데이터를 추가해 보자.(1시간 공부할 경우, 10분간 쉬는 시간을 제공해 보자.)

- '테이블 불러오기' 화면으로 들어가 'Train Data' 시트를 클릭한다. E열을 마우스 오른쪽 클릭하고, '오른쪽에 열 추가하기' 버튼을 클릭한다.

	A	B	C	D	E	F
1	날짜	요일	공부한 시간	시험 점수	공부를 시작한 시간	원쪽에 열 추가하기
2	2023-03-02	목	2	55	8:00:00	오른쪽에 열 추가하기
3	2023-03-03	금	3	69	8:03:00	열 삭제
4	2022-12-28	수	5	74	8:10:00	
5	2022-11-28	월	7	83	8:05:00	
6	2022-11-01	화	8	95	8:13:00	

- 새로운 열에 '제목'을 쉬는 시간, '데이터'를 0으로 입력한다.

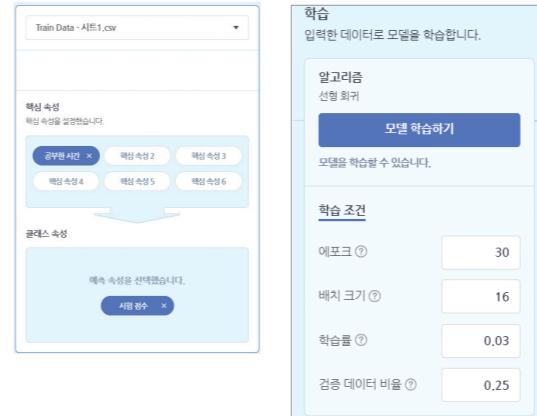
	A	B	C	D	E	F
1	날짜	요일	공부한 시간	시험 점수	공부를 시작한 시간	쉬는 시간
2	2023-03-02	목	2	55	8:00:00	0
3	2023-03-03	금	3	69	8:03:00	0
4	2022-12-28	수	5	74	8:10:00	0
5	2022-11-28	월	7	83	8:05:00	0
6	2022-11-01	화	8	95	8:13:00	0

## ② 데이터 전처리하기

- 다음과 같이 코드를 구성하여 수정된 테이블을 확인한다.



- x축에 놓을 요소를 '핵심 속성'에 드래그하고, y축에 놓을 요소를 '클래스 속성'에 드래그한다. 학습 조건을 설정하고, '모델 학습하기'를 클릭한다.



## 3. (지도학습)선형회귀를 이용하여 데이터 예측하기

### ① 데이터 학습하기

**Train Data**  
학습을 위한 데이터

**Test Data**  
예측하고 싶은 핵심 속성이 포함되어 있는 데이터

'Train Data'를 보면 공부한 시간이 길수록 시험 점수가 높은 것을 알 수 있다. 선형 회귀 모델을 이용하여 'Train Data'를 학습하고, 이를 바탕으로 'Test Data'의 시험 점수를 예측해 보자.

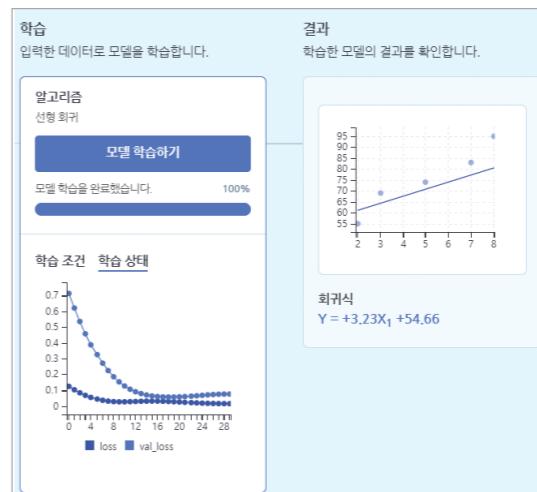
- 엔트리 메인 화면 중 '인공지능'을 클릭하고, '인공지능 모델 학습하기'를 클릭한다.
- 다양한 인공지능 모델 중 '예측:숫자(선형 회귀)'를 선택하고, '학습하기'를 클릭한다.



- 테이블 선택에서 'Train Data'를 선택한다.



- 학습 결과를 확인하고, '적용하기'를 클릭한다.



### ② 데이터 예측하기

- 다음과 같이 코드를 구성하여 'Test Data' 속에 있는 학습한 시간에 대한 시험 점수 예측값을 확인한다.



**에포크**  
입력한 데이터 전체를 몇 번 반복하여 학습할지 정하는 부분

**배치 크기**  
입력한 데이터 전체를 얼마큼 작은 부분으로 쪼개서 학습할지 정하는 부분

**학습률**  
데이터를 얼마나 세세하게 학습할지 정하는 부분

**검증 데이터 비율**  
어느 정도 비율을 학습한 모델을 검증하는 데에 사용할지 정하는 부분

학습 결과, 기울기가 음이 나올 수 있습니다. 그 이유는 데이터의 양이 적기 때문입니다. 학습하는 데이터의 양이 많아질수록 학습 결과를 바탕으로 한 예측값은 정확해집니다.

## 4. (지도학습)이미지 데이터 분류하기

포켓몬스터 이미지 파일  
다운로드 주소  
<https://zrr.kr/ZLPn>

포켓몬스터 이미지 출처  
<https://zrr.kr/KbXi>

### ① 이미지 데이터 학습하기

포켓몬스터의 캐릭터 하나하나에는 독특한 특징을 가지고 있다. 캐릭터는 일정한 경험치가 쌓이면 진화를 하게 된다. 어린 시절에는 진화 과정을 모두 외우고 있었는데 시간이 지나니 모두 잊게 되었다. 그렇다면 우리의 만물박사 인공지능에게 물어 봅시다. 엔트리를 이용하여 이미지 데이터를 학습시키고, 유사 이미지 찾기를 해 보자.

- 엔트리 메인 화면 중 '인공지능'을 클릭하고, '인공지능 모델 학습하기'를 클릭한다.



- 인공지능 모델 중 '분류:이미지'를 선택하고, '학습하기'를 클릭한다.



- 모델의 이름을 설정하고, 각 클래스에 데이터를 입력한다. 이미지 데이터를 입력하는 방법은 두 가지(업로드, 촬영)이다.



- '클래스 추가하기'를 통해 4개의 클래스를 만들고, 포켓몬 이미지 데이터를 각각 업로드한다.



출처 : 닌텐도



출처 : 닌텐도

- 업로드가 완료되면 '모델 학습하기'를 클릭하여 모델을 학습한다. 학습이 완료되면 다른 이미지를 업로드하여 학습결과를 확인할 수 있다. '적용하기'를 클릭한다.

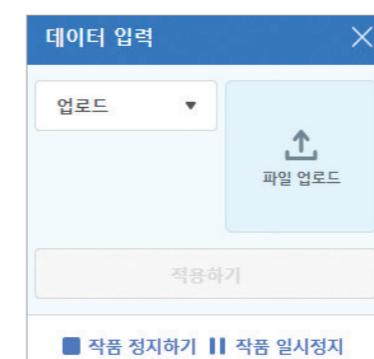


### ② 코드 구성하기



### ③ 유사 이미지 찾기

- 엔트리 메인 화면에서 '시작하기'를 클릭하면 데이터를 입력하라는 창이 뜬다. 다운로드 받은 이미지 중 '리자몽' 이미지를 업로드한다.



- 그 결과, 업로드한 '피카츄', '파이리', '이상해씨', '고복이' 중 가장 유사한 이미지와 신뢰도를 스프라이트가 말할 것이다. 그 결과를 통해 '리자몽'이 어떤 포켓몬의 진화형인지 알 수 있다.

## 5. (비지도학습) 군집 형성하기

(연습용) 군집화 데이터 다운로드  
<https://zrr.kr/69f0>

### ① 데이터 업로드 및 모델 학습하기

- ‘엔트리’ 메인 화면 중 ‘데이터 분석’을 클릭하고, ‘데이터 불러오기’를 클릭한다.(좌측의 연습용 데이터를 다운 받는다.)



데이터는 CSV 확장자로 다운로드 한다.

#### CSV란?

Creating Shared Value의 약자로 쉼표를 기준으로 항목을 구분하여 저장한 데이터

#### 테이블 선택하기

엔트리에서 수집해 놓은 데이터 활용

#### 파일 올리기

추가하고자하는 데이터를 업로드

#### 새로 만들기

빈 테이블을 추가

- ‘테이블 불러오기’ 화면 좌측에 ‘테이블 추가하기’ 버튼을 클릭한다.



- 상단에 ‘테이블 선택’, ‘파일 올리기’, ‘새로 만들기’ 버튼이 있다. 학습 상황에 맞는 탭을 선택한다.(본 내용에서는 ‘파일 올리기’를 선택한다.)



- 다운로드 받은 ‘k-means 군집화 데이터’를 업로드하고, ‘추가하기’ 버튼을 클릭한다.



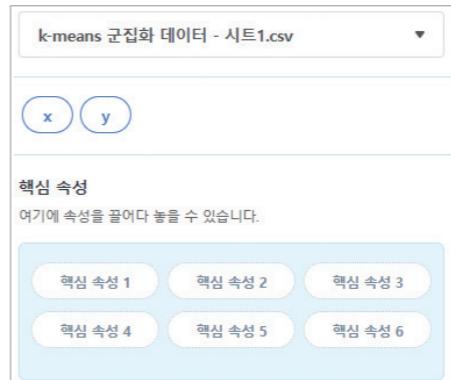
- 엔트리 메인 화면 중 ‘인공지능’을 클릭하고, ‘인공지능 모델 학습하기’를 클릭한다.



- 인공지능 모델 중 ‘군집:숫자(k-평균)’을 선택하고, ‘학습하기’를 클릭한다.



- 모델의 제목을 설정하고, ‘k-means 군집화 데이터’를 입력한다.



- ‘핵심 속성’을 x, y로 설정한다.



- 학습 조건을 설정하고, ‘모델 학습하기’를 클릭하여 입력한 데이터를 학습한다.



#### 학습 조건

군집의 개수는 대량적인 값으로 설정한다. 코딩을 통해 군집의 개수를 바꿔가며 최적의 군집을 설정한다.

- 학습 결과를 확인하고, '적용하기'를 클릭한다.

적용하기

## ② 코드 구성하기

군집의 개수를 설정하고, 군집의 중심점 좌표를 말하게 하는 코드를 구성한다.



## ③ 특정 좌표의 군집 찾기

특정 좌표가 어느 군집에 포함되는지 예측하는 활동이다. 아래와 같이 코드를 구성하면 특정 좌표의 군집을 예측할 수 있다.



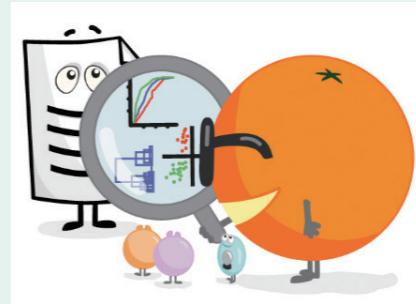
# ★오렌지3란?



- 오렌지3 사이트 주소  
<https://orangedatamining.com/>

## 1. 오렌지3 소개

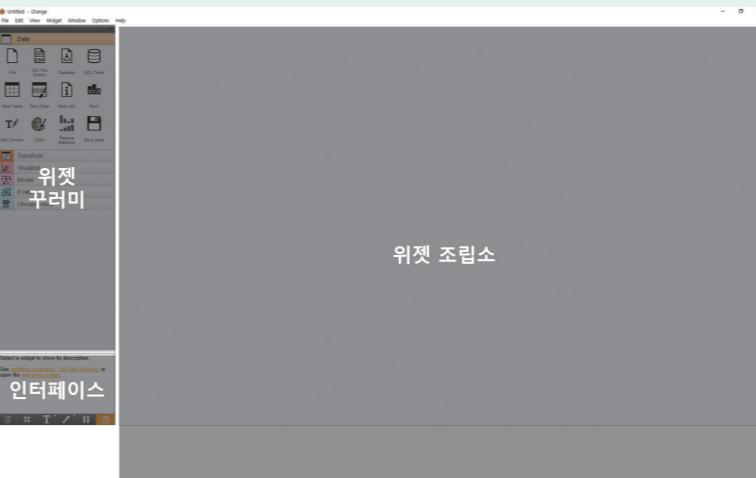
'오렌지3'는 데이터 분석 과정을 시각적·직관적으로 확인할 수 있는 프로그램이다. 통계 분포, 그림, 산점도를 탐색할 수 있고, 다양한 인공지능 모델을 활용할 수 있다.



## 2. 기본 사용법

### ● 화면 구성

'오렌지3'의 화면 구성은 '위젯 꾸러미', '위젯 조립소', '인터페이스'로 구성되어 있다.



**위젯 꾸러미**  
데이터 분석의 절차에 맞게 위젯이 구성

**위젯 조립소**  
'위젯 꾸러미'에서 위젯을 불러와 '위젯 조립소'에서 위젯을 연결

**인터페이스**  
'오렌지3'의 전체적인 활용은 '인터페이스' 및 상단 '메뉴'에서 조작 가능

# 02

## AI로 데이터를 시각화 및 분석해 보자(오렌지3)



오렌지3 사이트 주소  
<https://orangedatamining.com/>

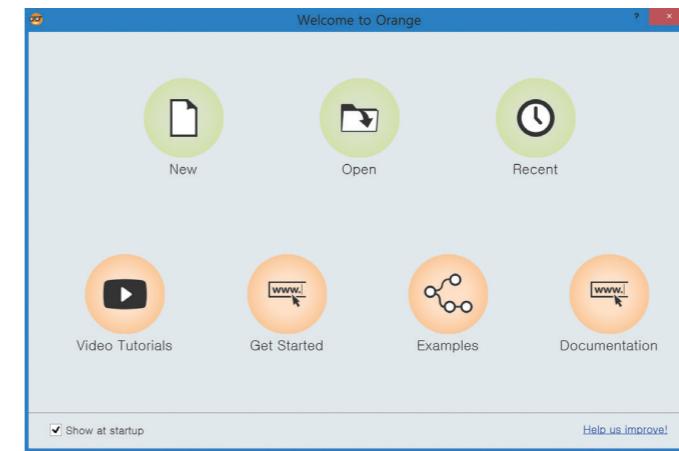
### 1. 데이터 업로드 및 확인하기

#### ① 데이터 업로드하기

- '오렌지3' 사이트에 접속하여 다운로드 받는다. 다운로드 후 설치 한다.



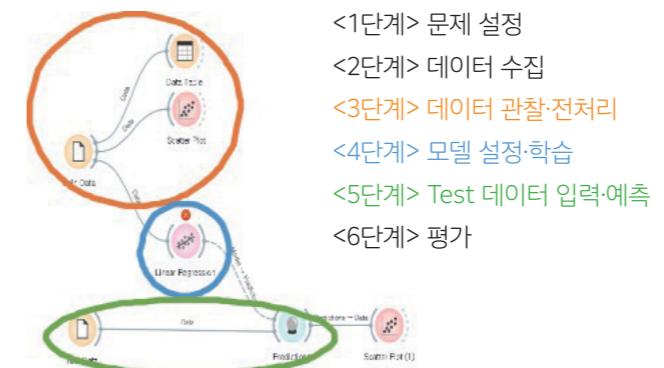
- '오렌지3'를 실행시키면 아래와 같은 화면이 나온다. 'NEW'를 클릭 하여 새로 시작한다.



- 화면 좌측에 여러 탭이 있다. 5가지 탭으로 구성되어 있으며, 데이터를 다룰 수 있는 다양한 위젯을 담고 있다.



- '오렌지3'를 이용하여 데이터를 분석하는 과정은 다음과 같다. 그 중 3, 4, 5단계를 '오렌지3'를 이용하여 분석한다.



(연습용) Train Data  
https://zrr.kr/yXas

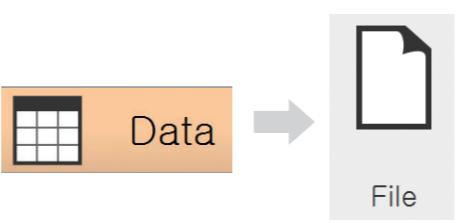
(연습용) Test Data  
https://zrr.kr/bPLT

데이터는 CSV 확장자로 다운로드 한다.

CSV란?  
Creating Shared Value의 약자로 쉼표를 기준으로 항목을 구분하여 저장한 데이터

파일을 업로드하는 방법도 있지만 URL을 이용하여 업로드하는 방법도 있다. 구글 스프레드시트를 이용한 데이터는 URL을 이용하는 것이 용이하다.

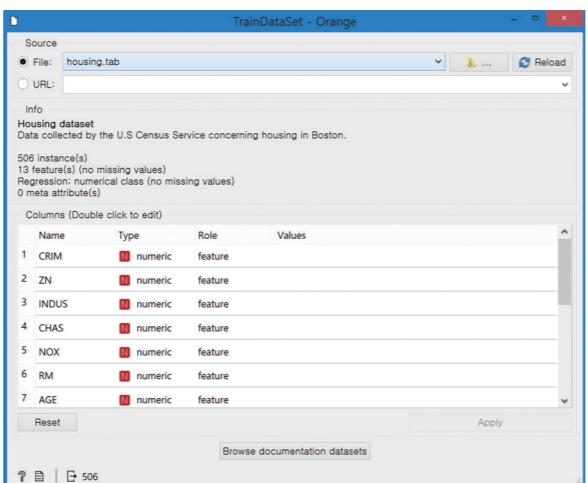
- 'Data' 탭을 선택하고, 'File'을 클릭한다.



- 추가된 'File'을 마우스 오른쪽 클릭하고, 'Rename'을 클릭하여 위젯의 이름을 바꾼다.



- 파일 위젯을 더블 클릭하면 아래와 같은 창이 뜬다. 상단에 파일을 클릭하여 업로드하고자 하는 데이터를 업로드한다.



- 업로드한 데이터를 확인하고, Type과 Role을 설정한다. 설정이 끝나면 'Apply'를 누른다.

Name	Type	Role
1 날짜	T datetime	skip
2 공부한 시간	N numeric	feature
3 시험 점수	N numeric	target
4 요일	S text	skip
5 공부를 시작한 ...	C categorical	skip

## ② 데이터 확인하기

- 'Data' 탭에서 'Data Table'을 불러온 후, 업로드한 데이터와 연결한다.

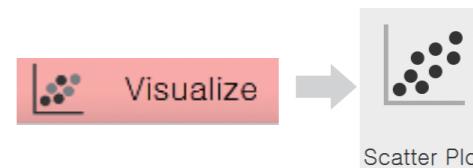


- 데이터를 확인한다.

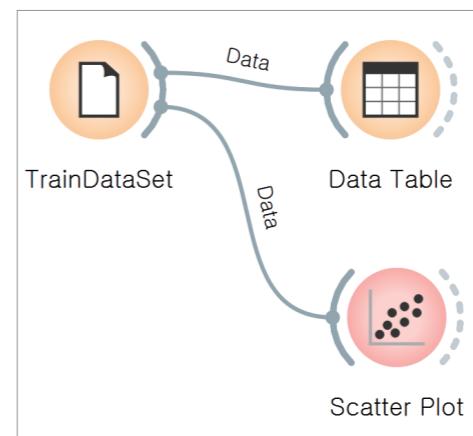
	시험 점수	공부한 시간
1	55	2
2	69	3
3	74	5
4	83	7
5	95	8

## ③ 데이터를 차트로 확인하기

- 'Visualize' 탭을 선택하고, 'Scatter Plot'을 불러옵니다.



- 'Visualize' 탭 중 'Scatter Plot'을 불러와 데이터와 연결한다.



'공부한 시간'(독립 변수)에 따른 '시험 점수'(종속 변수)를 알고자 하므로 '공부한 시간'은 Target, 공부한 시간은 Feature, 나머지 데이터는 'skip'으로 설정한다.

Feature Data  
종속 변수를 계산하기 위해 사용되는 독립 변수

Meta Data  
데이터에 관한 구조화된 데이터로, 다른 데이터를 설명해 주는 데이터

Target Data  
독립 변수에 영향을 받아서 값이 변화하는 변수

Skip  
건너뛰는 데이터

x축과 y축을 지정하고, 데이터의 크기에 따라 점의 색, 모양, 크기를 변경할 수 있다.

#### Show legend

범례 보기

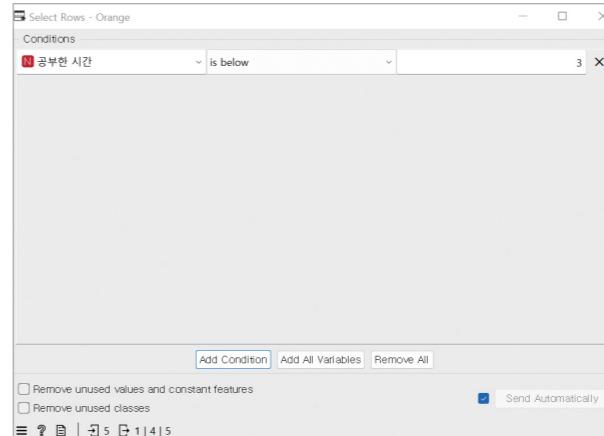
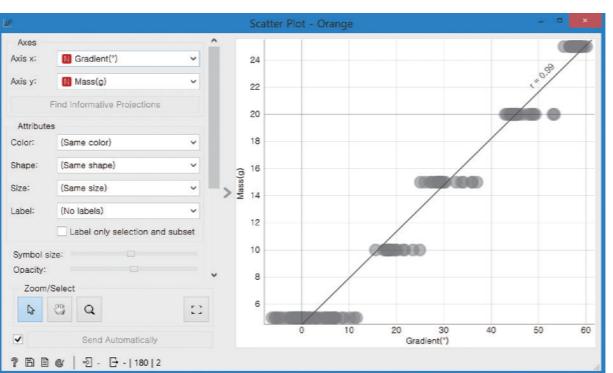
#### Show gridlines

그래프의 종선과 횡선

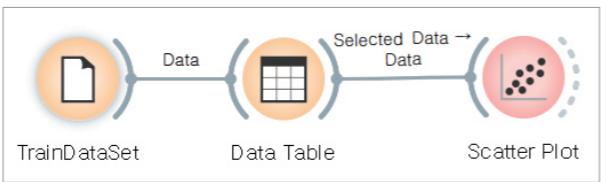
#### Show regression line

추세선 보기

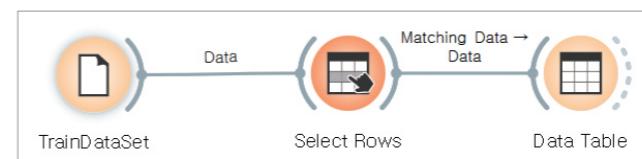
- 'Scatter Plot'을 더블 클릭하여 차트를 확인한다.



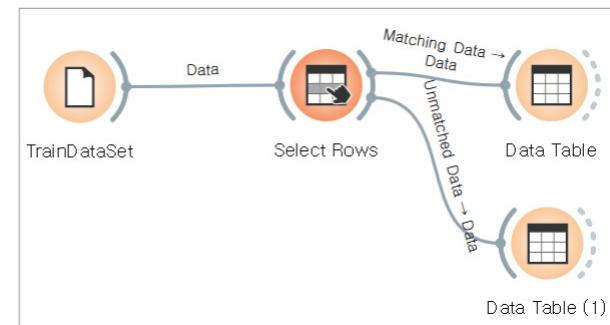
- 아래의 그림처럼 위젯을 연결해도 그래프를 확인할 수 있다.



- 'Data Table'과 연결하여 결과를 확인한다.



- '공부한 시간'이 3시간 이상인 데이터도 확인할 수 있다. 'Data Table'을 불러와 연결하고, 연결선 위에 적힌 정보를 'Unmatched Data → Data'로 수정한다.



Unmatched Data를 바꾸기 위해 연결선 위에 적힌 정보를 클릭하면 수정 창이 나온다. 다음과 같이 수정하면 된다.



## 2. 데이터 전처리하기

### 데이터 전처리란?

데이터 전처리란 모델의 성능을 높이기 위해 모델 투입 전에 수집한 자료를 제거하거나 조작하는 작업을 말한다.

#### ① '행' 선택하기

'공부한 시간' 데이터 중 필요한 데이터만 선별할 수 있다. 3시간 미만인 데이터만 선별해 보자.

- 'Transform' 탭에서 'Select Rows' 위젯을 불러와 Train Data와 연결한다.



- 'Select Rows'를 더블 클릭하고 '공부한 시간', 'is below', '3'을 선택한다.

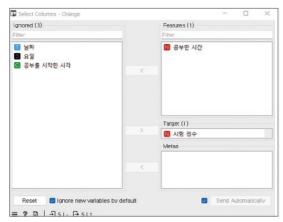
#### ② '열' 선택하기

데이터 분석에 필요한 열은 선택, 추가할 수 있다.

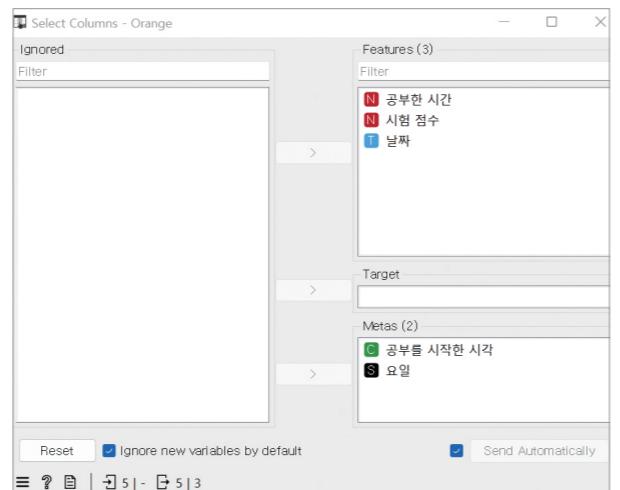
- 'Transform' 탭에서 'Select Columns' 위젯을 불러와 'Train Data'와 연결한다.



Unmatched Data를 바꾸기 위해 연결선 위에 적힌 정보를 클릭하면 수정 창이 나온다. 다음과 같이 수정하면 된다.



- 마우스를 드래그하거나 중앙 화살표를 이용하여 필요한 데이터를 선택한다.



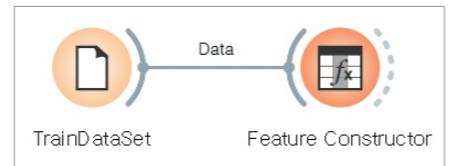
- 'Data Table'과 연결하여 결과를 확인한다.



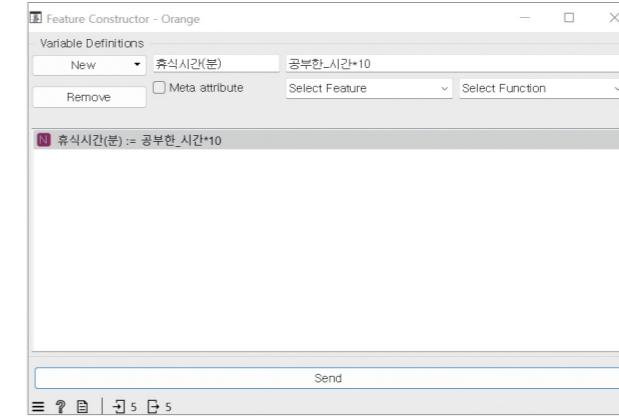
### ③ 데이터 생성하기

1시간 공부하면 10분 쉬는 시간을 제공하고자 한다. 공부한 시간에 따른 쉬는 시간 데이터를 추가하는 방법이다.

- 'Transform' 탭에서 'Feature Constructor' 위젯을 불러와 'Train Data'와 연결한다.



- 'Feature Constructor' 위젯을 더블 클릭하고, 다음과 같이 수정한 후 'Send' 버튼을 누른다.



- 'Data Table'과 연결하여 결과를 확인한다.

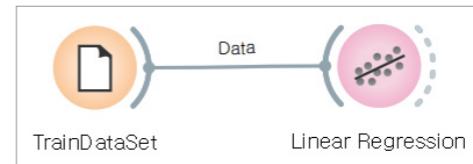


## 3. (지도학습)선형회귀를 이용하여 데이터 예측하기

### ① 데이터 학습하기

'Train Data'를 보면 공부한 시간이 길수록 시험 점수가 높은 것을 알 수 있다. 선형 회귀 모델을 이용하여 'Train Data'를 학습하고, 이를 바탕으로 'Test Data'의 시험 점수를 예측해 보자.

- 'Model' 탭에서 'Linear Regression'을 불러와 'Train Data'와 연결 한다.



**Train Data**  
학습을 위한 데이터

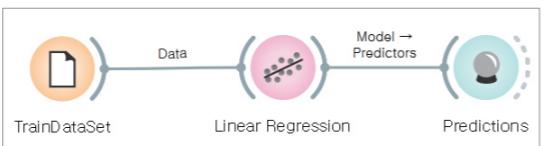
**Test Data**  
예측하고 싶은 핵심 속성이 포함되어 있는 데이터

## ② 데이터 예측하기

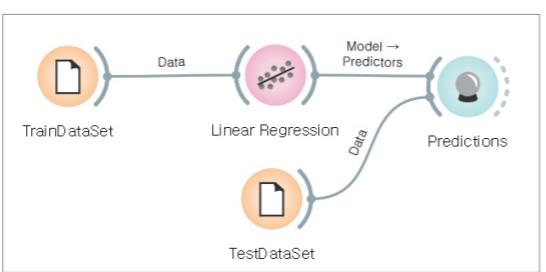
- 'Data' 탭에서 'File' 위젯을 불러와 Test Data를 업로드한다.



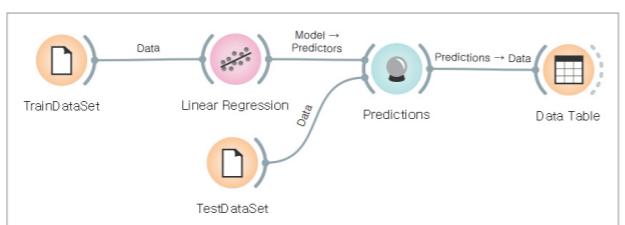
- 'Evaluate' 탭에서 'Predictions' 위젯을 불러와 'Linear Regression' 위젯과 연결한다.



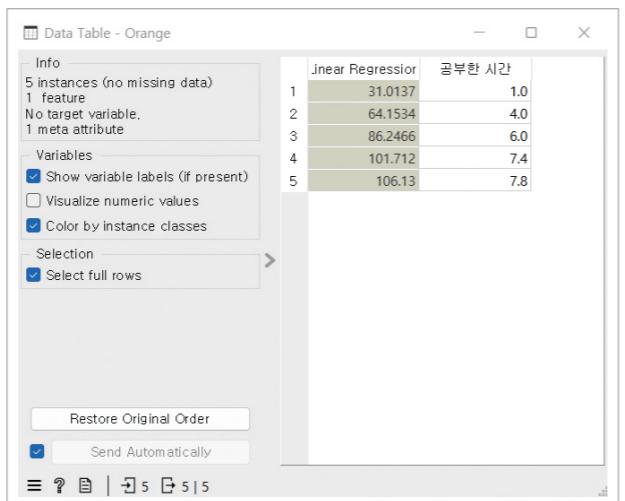
- Test Data와 'Predictions'를 연결한다.



- 'Data' 탭에서 'Data Table' 위젯을 불러와 'Predictions'를 연결 한다.



- 'Data Table' 위젯을 더블 클릭하여 예측한 결과를 확인한다.

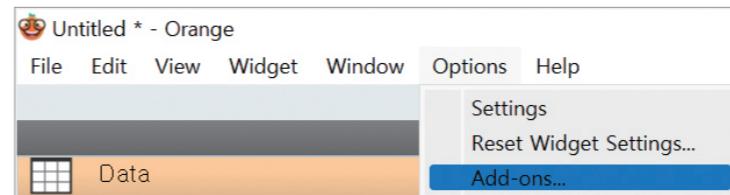


## 4. (지도학습) 이미지 데이터 분류하기

### ① '이미지 분석 확장' 탭 설치하기

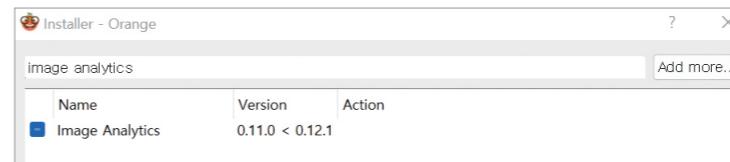
포켓몬스터의 캐릭터 하나하나에는 독특한 특징을 가지고 있다. 캐릭터는 일정한 경험치가 쌓이면 진화를 하게 된다. 어린 시절에는 진화 과정을 모두 외우고 있었는데 시간이 지나니 모두 잊게 되었다. 그렇다면 우리의 만물박사 인공지능에게 물어 봅시다. 오렌지3를 이용하여 이미지 데이터를 학습시키고, 유사 이미지 찾기를 해 보자.

- 'Options' - 'Add-ons...'를 클릭한다.

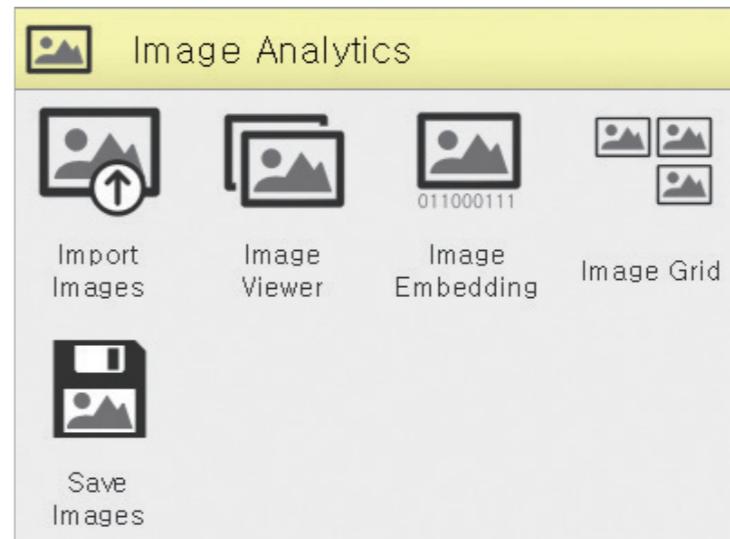


설치가 완료되면 오렌지3를 종료하고, 재실행한다.

- 'Image Analytics'를 검색하고, 설치한다.



- 'Image Analytics' 탭이 추가되었는지 확인한다.



## ② 이미지 분류하기

포켓몬스터 이미지 파일  
다운로드 주소  
<https://zrr.kr/ZLPn>

포켓몬스터 이미지 출처  
<https://zrr.kr/KbXi>

**Embedding**  
숫자로 되어 있지 않은 자료를 숫자와 같은 수치적 형태로 매핑하는 작업

이번 연습에서 사용하는 이미지는 'VGG-16'를 이용하였을 때 좋은 결과를 보이므로 'VGG-16'을 선택한다.

- 'Image Analytics' 탭에서 'Import Images'를 불러와 이미지를 업로드한다.

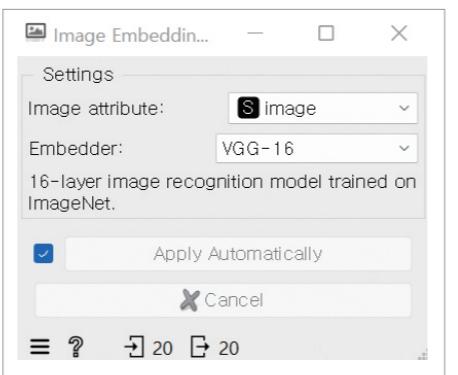


(예시 이미지 20장(피카츄, 파이리, 이상해씨, 꼬북이)을 추가한다.)

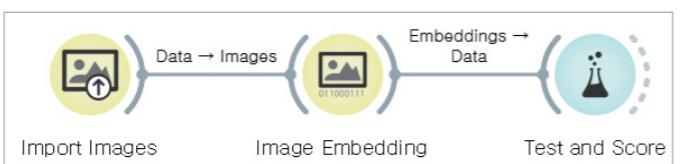
- 'Image Analytics' 탭에서 'Image Embedding' 위젯을 불러와 연결한다.



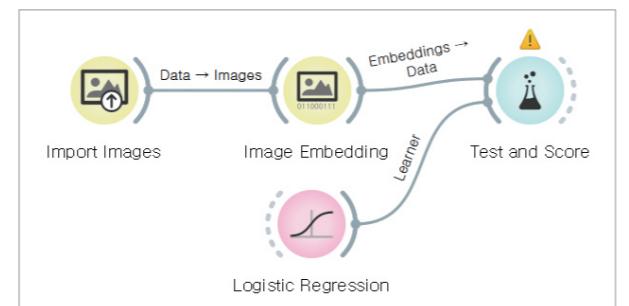
- 'Image Embedding' 위젯을 더블 클릭하고, 분류하고자 하는 이미지에 적합한 'Embedder'를 선택한다.



- 'Evaluate' 탭에서 'Test and Score' 위젯을 불러와 'Image Embedding'와 연결한다.



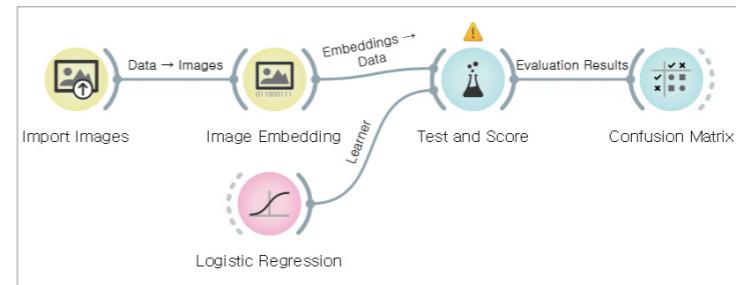
- 'Model' 탭에서 'Logistic Regression' 위젯을 불러와 'Test and Score'와 연결한다.



**Test and Score**  
Train Data와 학습알고리즘을 연결해서 성능을 평가할 때 사용하는 위젯

**Logistic Regression**  
예측 분석을 위한 회귀분석 중에서 특히 종속 변수가 이분형일 때 수행 할 수 있는 회귀 분석 기법의 한 종류  
(데이터가 어떤 범주에 속할 확률을 0에서 1 사이의 값으로 예측한 뒤 그 확률에 따라 가능성이 특정 기준치 이상인 경우 해당 클래스로 분류해 주는 지도학습 알고리즘)

- 'Evaluate' 탭에서 'Confusion Matrix' 위젯을 불러와 'Test and Score'와 연결한다.



- 'Image Viewer'를 'Confusion Matrix'과 연결한 후 두 위젯을 더블 클릭하여 창을 띄운다. 각 데이터를 클릭하면 어떤 이미지가 예측과 실제가 일치하는지 일치하지 않는지 확인한다.

		Predicted				
		이미지 분류#꼬북이	이미지 분류#이상해씨	이미지 분류#파이리	이미지 분류#피카츄	
Actual	이미지 분류#꼬북이	4	0	1	0	5
	이미지 분류#이상해씨	2	3	0	0	5
Actual	이미지 분류#파이리	0	1	4	0	5
	이미지 분류#피카츄	0	1	0	4	5
		Σ	6	5	5	20

## ③ 유사 이미지 찾기

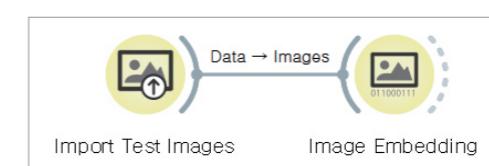
리자몽은 파이리의 진화형 캐릭터이다. 피카츄, 이상해씨, 꼬북이 캐릭터보다 모습이 가장 유사하다. 이미지의 유사도를 파악하여 리자몽이 파이리의 진화형임을 찾아 낼 수 있다.

'Image Analytics' 탭 추가는 이전에 설명한 '이미지 분석 확장' 탭 설치하기를 참고한다

- 'Image Analytics' 탭에서 'Import Images' 위젯을 불러온 후 위젯 이름을 'Import Test Images'로 변경한다. 그 후 리자몽 이미지를 업로드한다.



- 'Image Embedding'을 업로드한 이미지와 연결하여, Embedding 한다.



이번 연습에서 사용하는 이미지는 'VGG-16'을 이용하였을 때 좋은 결과를 보이므로 'VGG-16'을 선택한다.

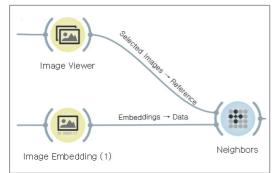
- 'Image Viewer'를 'Image Embedding'와 연결하여 이미지가 잘 업로드되었는지 확인한다.



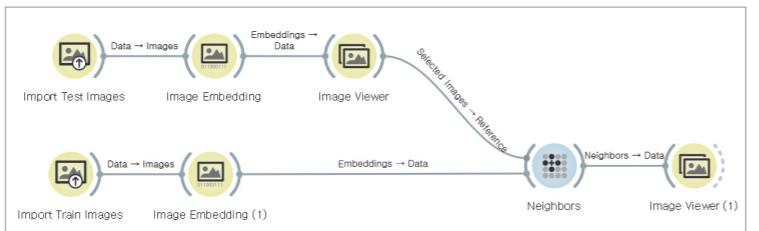
- 'Image Analytics' 탭에서 'Import Images' 위젯을 불러온 후 위젯 이름을 'Import Train Images'로 변경한다. 그 후 각 캐릭터의 이미지를 업로드한다.



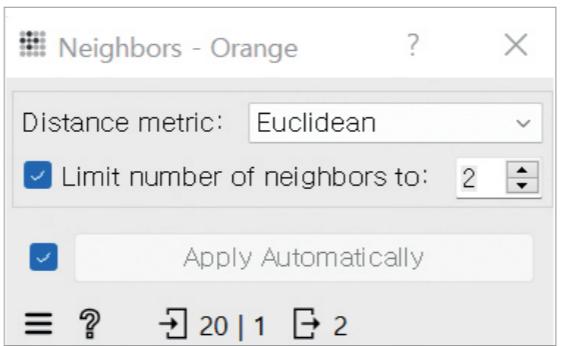
'Neighbors' 위젯과 연결된 선 위의 설정이 중요하다. Test Image는 Reference로, Train Image는 Data로 설정한다.



- 유사 이미지를 찾기 위해 'Unsupervised' 탭에서 'Neighbors' 위젯을 불러와 업로드 시킨 두 이미지 그룹을 연결한다.



- 'Neighbors'를 더블 클릭한 후, Limit number of neighbors to를 설정한다. 유사한 이미지를 몇 개 보일 것인지 설정할 수 있다.

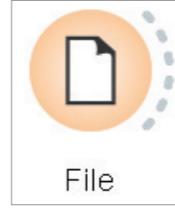


- 'Neighbors' 위젯과 연결된 'Image Viewer'를 보면 리자몽과 유사한 이미지가 무엇인지 알 수 있다.

## 5. (비지도학습) 군집 형성하기

### ① 데이터 업로드 및 모델 학습하기

- 'Data' 탭에서 'File' 위젯을 불러온 후, 'k-means 군집화 데이터' 파일을 업로드한다.

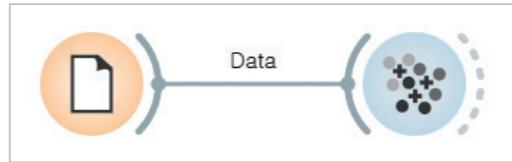


(연습용) 군집화 데이터 다운로드  
<https://zrr.kr/69f0>

데이터는 CSV 확장자로 다운로드 한다.

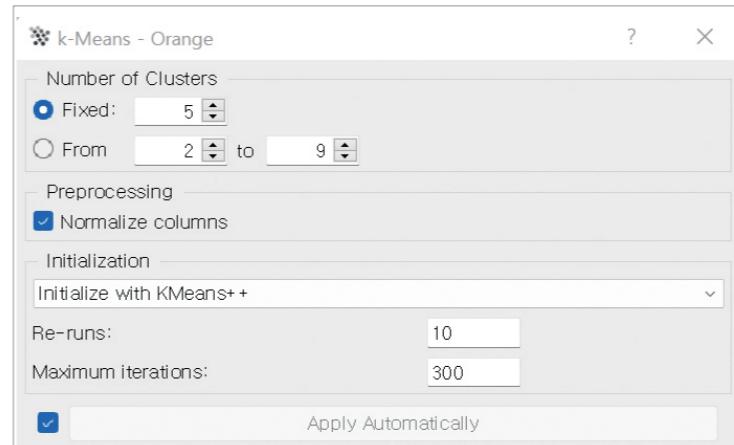
**CSV란?**  
Creating Shared Value의 약자로 쉼표를 기준으로 항목을 구분하여 저장 한 데이터

- 'Unsupervised' 탭에서 'K-Means' 위젯을 불러와 'File'과 연결 한다.



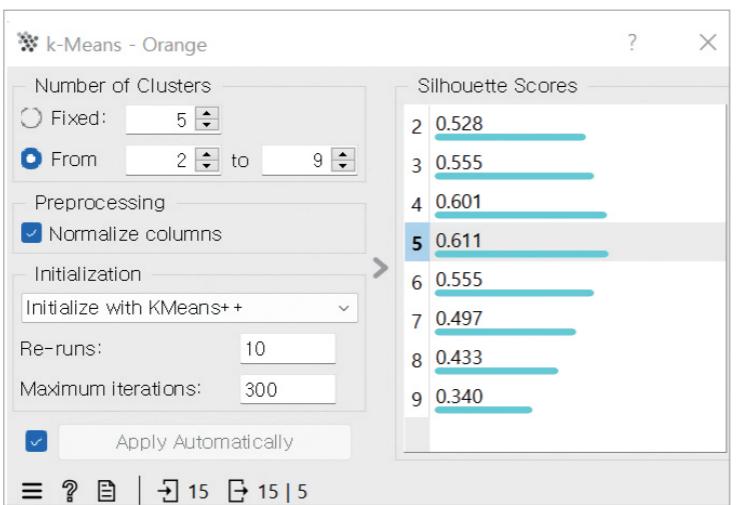
### ② 군집 설정하기

- 'K-Means' 위젯을 더블 클릭하면 군집의 개수를 설정할 수 있다. 'Fixed'를 설정하면 군집의 개수를 지정할 수 있고, 'From'을 설정하면 군집의 개수를 범위로 지정할 수 있다.



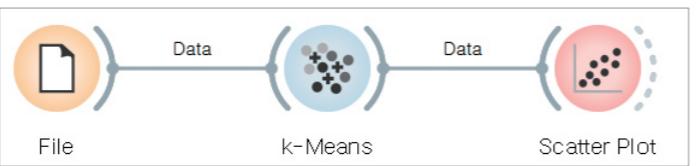
- 군집의 개수를 범위로 설정할 경우, 가장 적합한 군집의 개수를 'Silhouette Scores'를 통해 확인할 수 있다.

연습용 데이터에서는 5개의 군집이 가장 적합한 것을 알 수 있다.

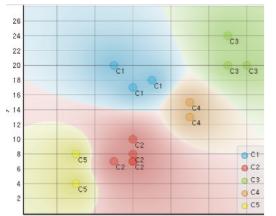


### ③ 군집 확인하기

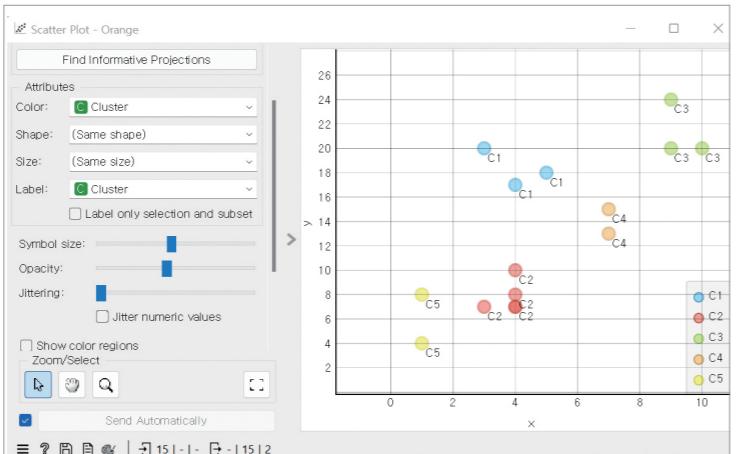
- 'Visualize' 탭에서 'Scatter Plot' 위젯을 불러온 후, 'K-means' 위젯과 연결한다.



Show Color regions를 체크하면 군집의 영역을 색으로 확인할 수 있다.



- 'Color'와 'Label'을 Cluster로 설정하면 군집을 색과 라벨로 구별 할 수 있다.



## II. AI로 응용프로그램 개발

이 장에서는 스크래치 프로그래밍 환경을 넘어서, 고급 인공지능 기술의 하나인 머신러닝을 탐구한다. 스크래치에서 사용 가능한 머신러닝 블록을 통해, 복잡한 문제들을 해결하는 혁신적인 방법을 배우게 된다. 기초적인 스크래치 사용법에 익숙해진 후, 이번 단원에서는 머신러닝의 기본 원리를 이해하고, 다양한 활용 사례를 살펴보면서 실제 프로그래밍 프로젝트에 어떻게 적용할 수 있는지 배울 기회를 갖도록 한다.

### 1. ML2스크래치

### 2. 핸즈 포즈

# 01

## ML2 스크래치



Q. 스크래치로 인공지능을 어떻게 활용할까?

Q. 머신러닝으로 만들 수 있는 것이 무엇이 있는가?

### 1. 스크래치

"스크래치"는 어린이와 청소년을 위해 만들어진 프로그래밍 도구이다. MIT 미디어 랩에서 만들어졌으며, 블록 형태의 간단한 명령어를 조합하여 프로그래밍을 배울 수 있다. 이런 스타일의 프로그래밍 도구는 스크래치 외에도 '엔트리'나 '엠블록' 같은 다른 도구들도 있으며, 이런 도구들은 프로그래밍을 처음 접하는 학생들이 쉽게 이해하고 배울 수 있도록 설계된다.



그림 6\_블록기반언어

#### ① 스크래치의 특성

- 시각적 프로그래밍 언어: 스크래치는 코드를 작성하는 대신에 색상별로 구분된 블록들을 드래그 앤 드롭하여 프로그램을 만들 수

있다. 이를 통해 초보자도 직관적으로 프로그래밍의 원리를 이해할 수 있다.

- 예로 비슷한 방법으로 사용하는 블록은 같은 색으로 구성되어 있다.
- 교육적 목적: 스크래치는 프로그래밍 학습을 목적으로 설계되어 있다. 따라서 학습자가 쉽게 접근하고 실험할 수 있게 만들어져 있다.
- 다양한 프로젝트 생성: 애니메이션, 게임, 인터랙티브 스토리 등 다양한 프로젝트를 스크래치로 제작할 수 있다.
- 커뮤니티 기반: 스크래치 웹사이트는 사용자들이 자신의 프로젝트를 공유하고 다른 사람의 프로젝트를 볼 수 있는 커뮤니티 기능을 제공한다.
- 확장성: 스크래치는 다양한 하드웨어 플랫폼과 연동이 가능하므로, 예를 들어 로봇과 연동하여 로봇을 움직이게 하는 것도 가능하다.
- 국제적 지원: 스크래치는 전 세계 수많은 교육 기관과 커뮤니티에서 사용되고 있으며, 다양한 언어로 지원되고 있다.

#### ② 스크래치의 확장성

스크래치의 확장성은 그 학습 도구로서의 가치를 넘어서 다양한 환경과 상황에서 활용될 수 있고 확장된 관련 특징은 다음과 같다.

- 외부 하드웨어 연동: 디바이스(예: LEGO Mindstorms, micro:bit)와 연결하여 상호작용하는 프로그램 생성 가능
- 확장 블록: 스크래치 3.0에서 추가된 기능으로, 특정 하드웨어나 서비스에 특화된 프로그래밍이 가능
- 스크래치X: 웹 기반 API와 연동, 웹 서비스와 상호작용
- 플랫폼 지원: 웹 외의 다른 애플리케이션에서도 실행 가능
- 커뮤니티 기반 확장: 사용자들의 확장 기능 공유를 통한 기능 확장
- 다양한 주제 활용: 애니메이션, 게임 외에도 예술, 음악, 과학 등에서 활용 가능



그림 7\_블록언어 프로그램의 확장성

#### MIT 미디어 랩

1985년에 설립된 연구 기관으로, 기술과 디자인의 결합을 통한 혁신적 연구를 진행한다. 로보틱스, 인공지능, 가상 현실 등 다양한 분야에서 인간 중심의 기술을 탐구하며, 그 결과는 사회와 문화에 큰 영향

#### 인터랙티브

사용자의 행동이나 선택에 반응하는 시스템이나 미디어를 의미하며, 상호 작용을 가능하게 하는 특성

#### 하드웨어 플랫폼

특정 소프트웨어나 애플리케이션을 실행하기 위한 물리적인 기반 장치나 기술 구조를 의미

#### 확장성

시스템이 추가 블록이나 구성요소를 쉽게 포함하도록 하는 것

#### 웹 기반

API는 인터넷을 통해 서비스나 데이터를 제공받거나 요청하는 프로그래밍 인터페이스

#### 애플리케이션

특정 작업을 수행하기 위한 컴퓨터 프로그램 또는 모바일 소프트웨어

#### 커뮤니티

공통의 관심사나 목적을 가진 사람들이 형성하는 모임이나 네트워크를 의미

#### 플랫폼

비주얼 방식으로 코드를 생성할 수 있는 프로그래밍 환경을 의미

사이트는 <https://stretch3.github.io/>의 주소로 들어가면 된다.

## 2. 화면구성

### ① 스크래치 기본 화면

#### 백드롭

무대나 촬영 장소 뒤쪽에 위치한 배경 물품이나 화면을 가리키는 용어

#### 프로젝트

특정 목표를 달성하기 위해 계획적으로 수행하는 일련의 작업 및 활동

#### 스프라이트

독립적으로 움직일 수 있는 그래픽 객체나 캐릭터를 가르킴.

#### 객체

객체는 데이터와 그 데이터에 관련된 기능을 함께 갖는 독립적인 프로그래밍 단위

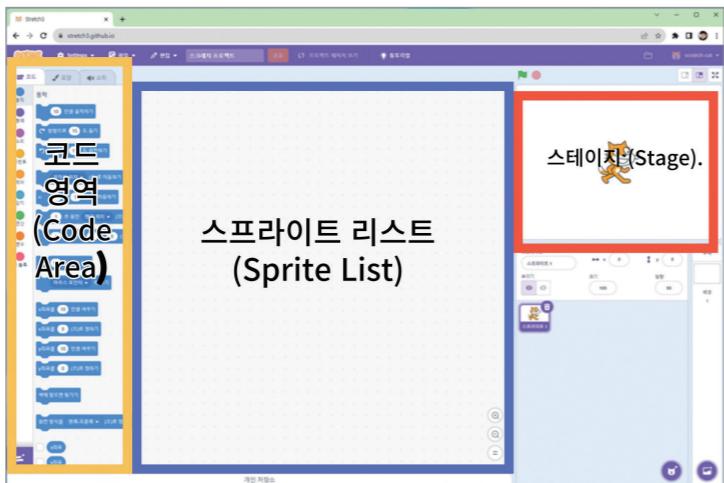


그림 8\_스크래치 기본화면

#### \* 스테이지 (Stage):

- 화면의 오른쪽 상단에 위치한다.
- 사용자의 프로그램이 실행되며 애니메이션, 상호작용 등을 실시간 가능
- 배경(백드롭)을 추가하거나 변경
- 전체 화면 모드로 전환하여 프로젝트를 전체 화면에서 실행

#### \* 스프라이트 리스트 (Sprite List):

- 화면의 중앙 아래쪽에 위치
- 프로젝트에 사용되는 모든 스프라이트(캐릭터나 객체)들의 리스트
- 새로운 스프라이트를 추가하거나, 기존 스프라이트를 수정하거나 삭제 가능.
- 각 스프라이트를 선택하면 해당 스프라이트의 코드를 볼 수 있게 되며, 코드 영역에서 프로그래밍 가능

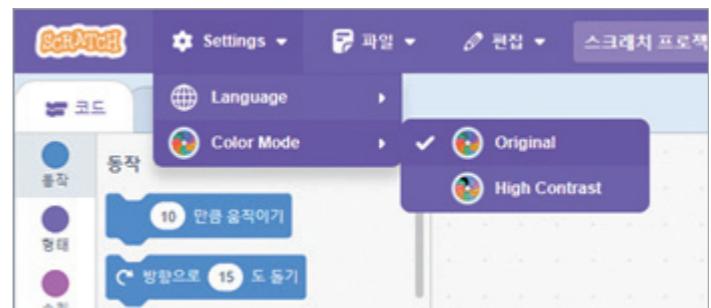
#### \* 코드 영역 (Code Area):

- 화면의 중앙에서 왼쪽에 걸쳐 위치

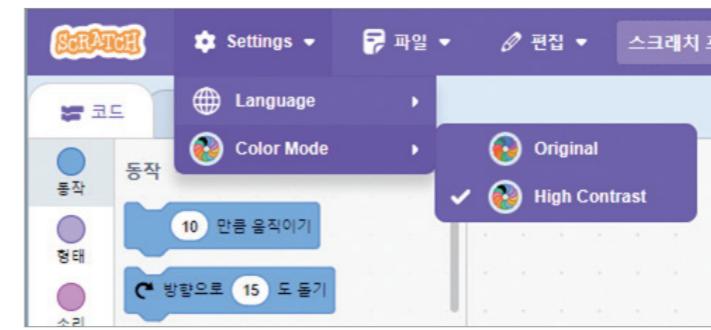
- 스프라이트에게 수행시킬 동작을 프로그래밍
- 블록들은 카테고리별로 구분되어 있으며, 원하는 블록을 드래그하여 스프라이트의 명령어를 조합
- 카테고리에는 동작, 모양, 소리, 이벤트, 제어, 감지, 연산, 변수, 내리스트
- 각 블록은 특정 기능을 수행하며, 블록들을 조합함으로써 복잡한 동작이나 로직을 구현

## 3. 머신러닝을 사용하기 위해 기본방법 익히기

### ① 블록모양 설정하기



블록모양 설정하기(오리지널)



블록모양 설정하기(High Contrast)

### ② 인공지능 확장 설정하기

- 스크래치에 인공지능 확장블록을 다운로드를 받기 위해 팔лет트창 아래 확장을 클릭한다.
- 아래와 같은 다양한 인공지능관련 확장 블록이 나온다.

#### 카테고리

프로그래밍에서 사용되는 블록 기반의 언어를 의미하며, 이러한 언어는 프로그래밍 명령어를 그래픽 블록으로 표현

#### 이벤트

프로그램이 반응해야 하는 특정한 상황이나 사용자의 행동

#### 머신러닝

컴퓨터가 데이터를 기반으로 학습하여 예측, 분류, 군집화 등의 작업을 수행하게 하는 알고리즘과 기술의 집합

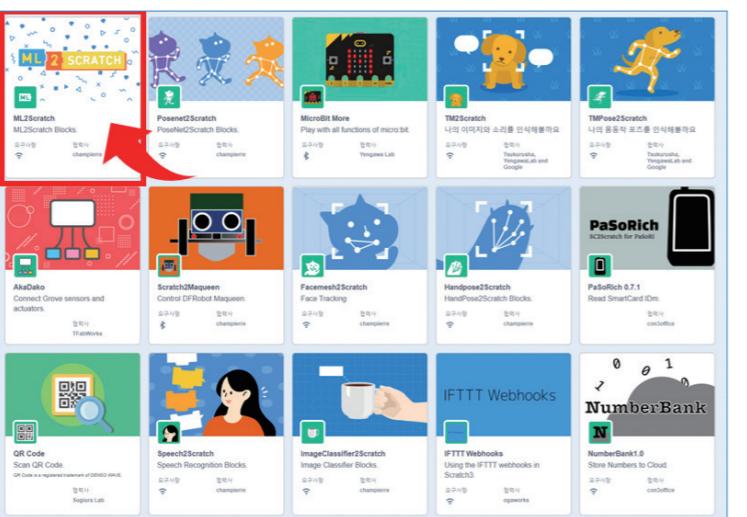
## 팔렛트창

사용 가능한 모든 블록들의 모음을 보여주는 영역, 블록들은 기능별로 카테고리로 분류되어 있음

## 팔렛트창의 사용도

하는 블록을 찾아 드래그하여 작업 영역에 가져다 놓음으로써 프로그램을 구성

- 원하는 인공지능 확장을 클릭하여 확장블록을 다운받는다.



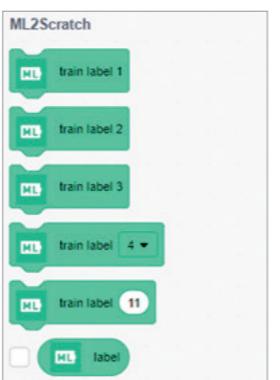
- 스크레치에 인공지능 확장블록을 다운로드를 받기 위해 팔렛트창 아래 확장을 클릭한다
- 아래와 같은 다양한 인공지능관련 확장 블록이 나온다.
- 원하는 인공지능 확장을 클릭하여 확장블록을 다운받는다.
- 여기서 “ML2SCREACH”를 선택한다.



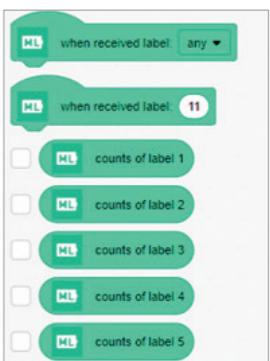
## ③ 추가한 블록 알기

머신러닝을 시킬 때 유의점  
배경은 항상 동일한 장소에서 실시 한다.  
조명이 밝은 곳에서 한다.

- 머신러닝 학습을 하기 위해 사용가능 블록
- 블록은 트레이닝 블록은 기본적으로 1~3번까지 사용
- 추가 머신러닝을 하기 위해서는 콤보박스 및 숫자를 기입



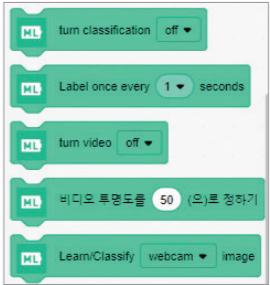
- 머신러닝으로 학습한 레이블을 선택 가능
- 해당 레이블 값을 활용



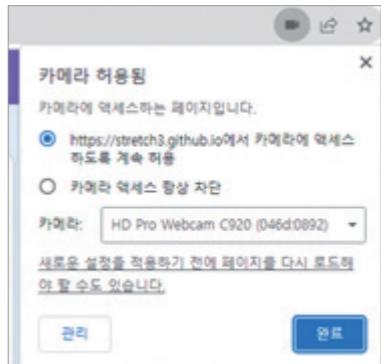
- 학습한 데이터를 리셋가능
- 머신러닝으로 학습한 데이터를 업로드 또는 다운로드 가능



- 비디오를 켜고, 끄기 가능
- 비디오의 투명도 조절
- 학습과 분류 이미지



## ④ 기초작업, 머신러닝 익히기

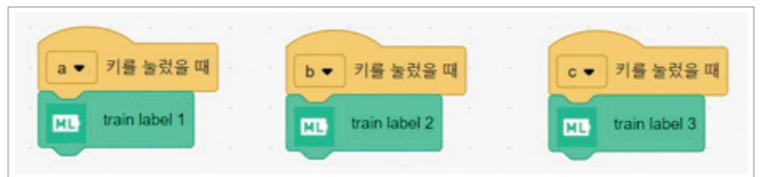


## 크롬 브라우저

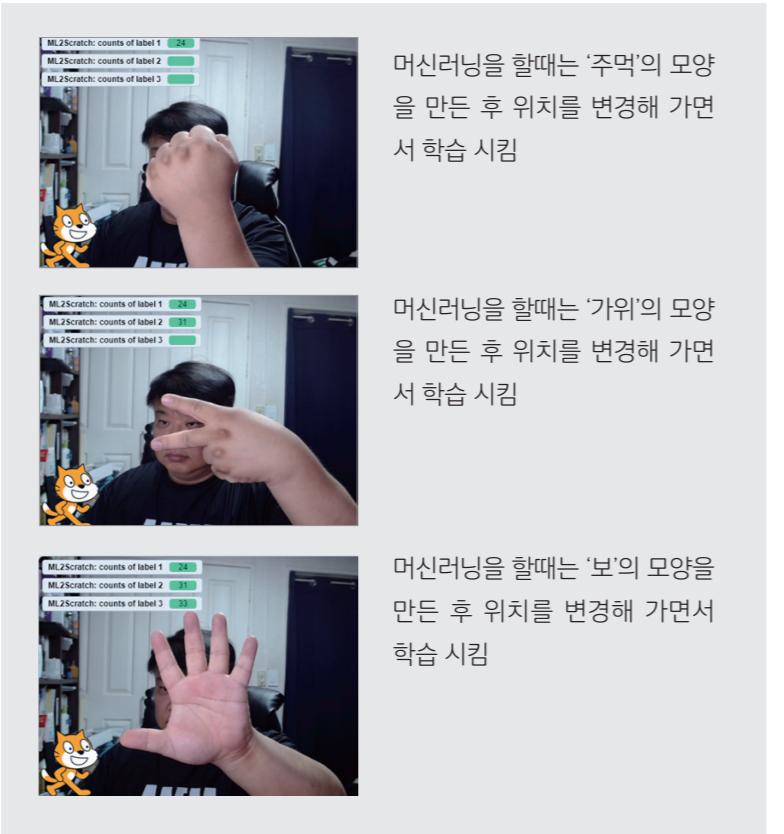
Google에서 개발한 무료 웹 브라우저로, 빠른 속도, 안정성, 확장성을 제공하며 다양한 플랫폼에서 사용 그라피 사용자 인터페이스와 웹 표준을 지원하며, 확장 프로그램을 통해 기능을 추가

- 인공지능 머신러닝 학습하기 전 2개 이상 카메라 사용할 경우
- 카메라 설정은 할 크롬브라우저 우측상단에 카메라 모양의 아이콘을 클릭
- 카메라를 적합하게 설정
- 그리고 카메라 액세스는 항상 허용

### 1) 머신러닝 학습시키기



- 인공지능 머신러닝을 하기 위해 왼쪽 블록과 같이 프로그래밍
- 'a' 키를 누르면 label 1에 머신러닝후 카운트 증가
- 'b' 키를 누르면 label 2에 머신러닝후 카운트 증가
- 'c' 키를 누르면 label 3에 머신러닝후 카운트 증가



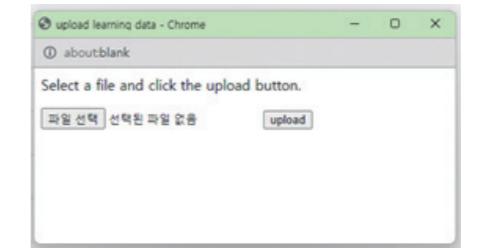
### 2) 학습한 머신러닝 파일만들기



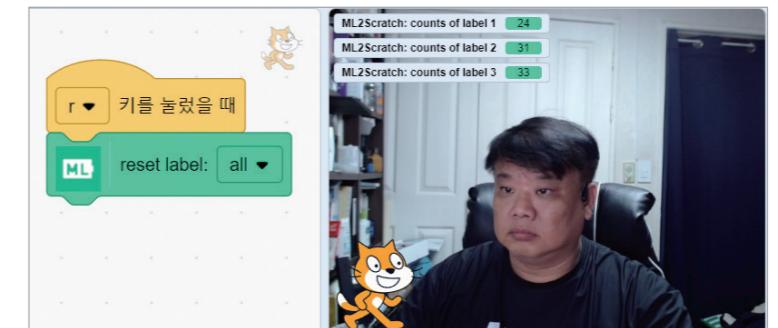
- 머신러닝 후 다운로드 파일 생성
- 오른쪽 그림과 같이 파일이 생성됨
- 일반적으로 download폴더 저장됨



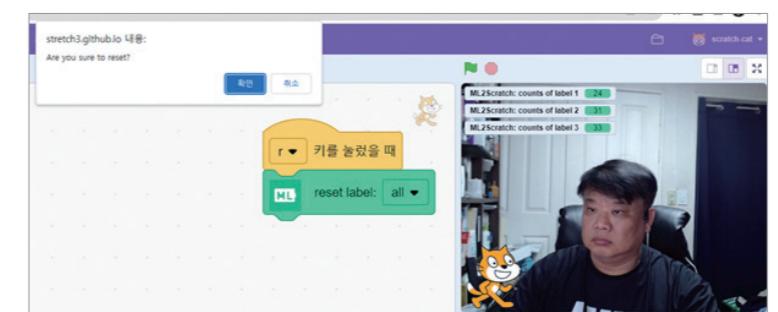
- 머신러닝 파일을 업로드 파일을 불러오기
- 파일선택을 누르고 파일을 찾은 후 업로드 클릭하면 기존의 저장된 파일을 사용할 수 있음.



### 3) 머신러닝 초기화 만들기



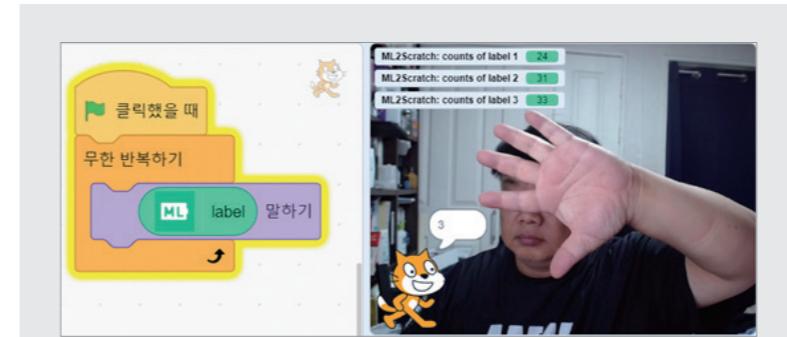
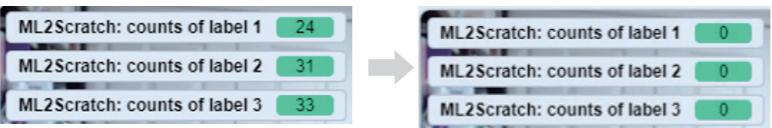
- 머신러닝후 lable을 초기상태로 만들 때
- 'r'키를 누르면 모든 lable을 초기화하기



- 'r'키를 누른 후 확인 버튼을 누르면 모든 label가 초기화됨



- 만약에 특정 label을 삭제하기 위해서는 콤보박스를 클릭하여 지우기 원하는 label을 선택
- label가 초기화
- 변수의 숫자를 확인 모든 label 카운트가 0으로 변경.

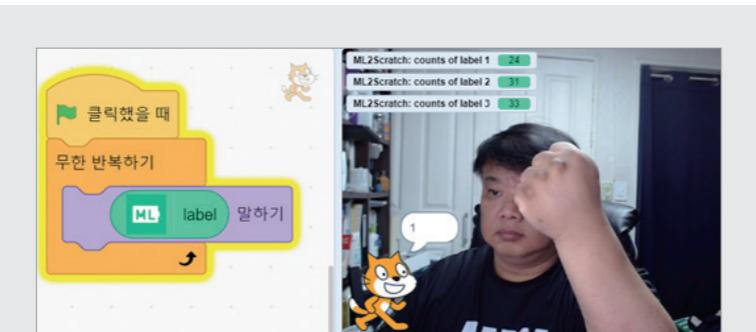


\* label 3은 '보'를 캐릭터가 3을 출력되는지를 확인한다.

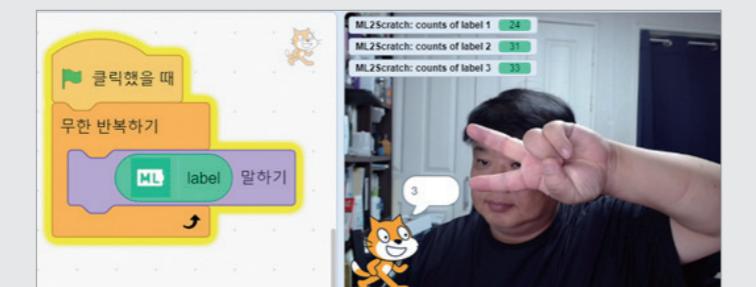
- 위 작업에서 확인을 되었다면 머신러닝으로 학습한 데이터를 활용하여 각 동작에 따라 캐릭터가 해당 label에 숫자를 말하도록 간단한 프로그램을 만들어보자.

## ⑤ 간단한 프로그램 만들기

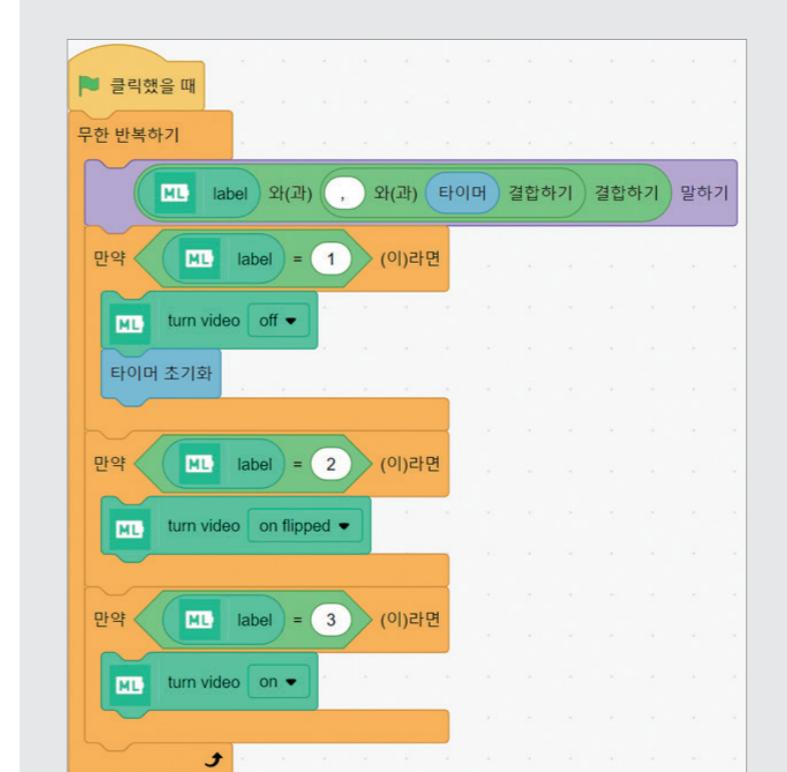
- 프로그램을 만들기 전 다음과 같은 블록으로 머신러닝의 값의 출력 값을 테스트를 한다.



\* label 1은 '주먹'을 캐릭터가 1을 출력되는지를 확인한다.



\* label 2은 '가위'를 캐릭터가 2를 출력되는지를 확인한다.



- '주먹'이 인식이 되면 카메라를 크고 타이머를 초기화 시키도록
- '가위'가 인식 되면 카메라가 화상 반전되도록 한다.
- '바위'가 인식 되면 카메라가 처음화면으로

위의 프로그램을 다행하게 진행하도록 한다.

## 주의점

프로그램을 개발할 때 머신러닝 모델의 학습 모델 학습이 미흡하다고 판단 될 때는, 더 많은 학습 데이터를 활용하거나 다양한 학습 전략을 적용해야 할 수도 있다.



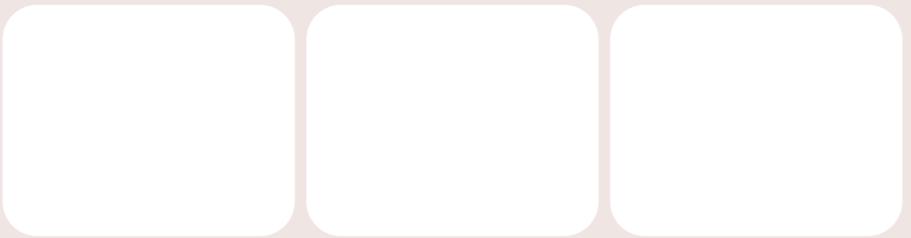
### 프로젝트 만들기

#### '인형을 춤추게 하는 인공지능'

##### 1) 기본아이디어

머신러닝으로 학습된 결과를 가지고 캐릭터 발레리나가 발레를 춤을 추는 작품을 만들어 보도록 하자. 그럼 아래 과정으로 만들어야 할지 생각해보자.

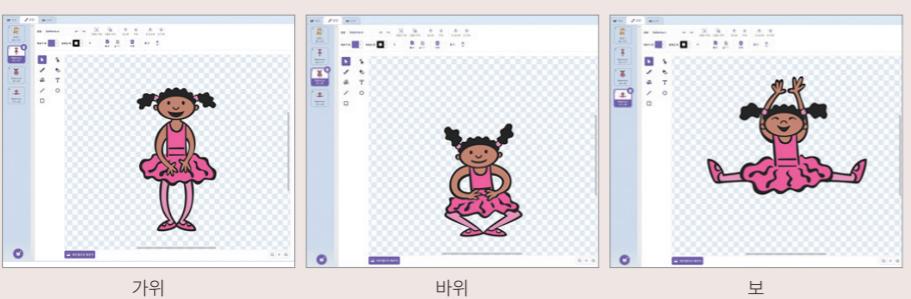
##### ▶ 1단계: 머신러닝 학습만들기(어떻게 할지 적어보세요)



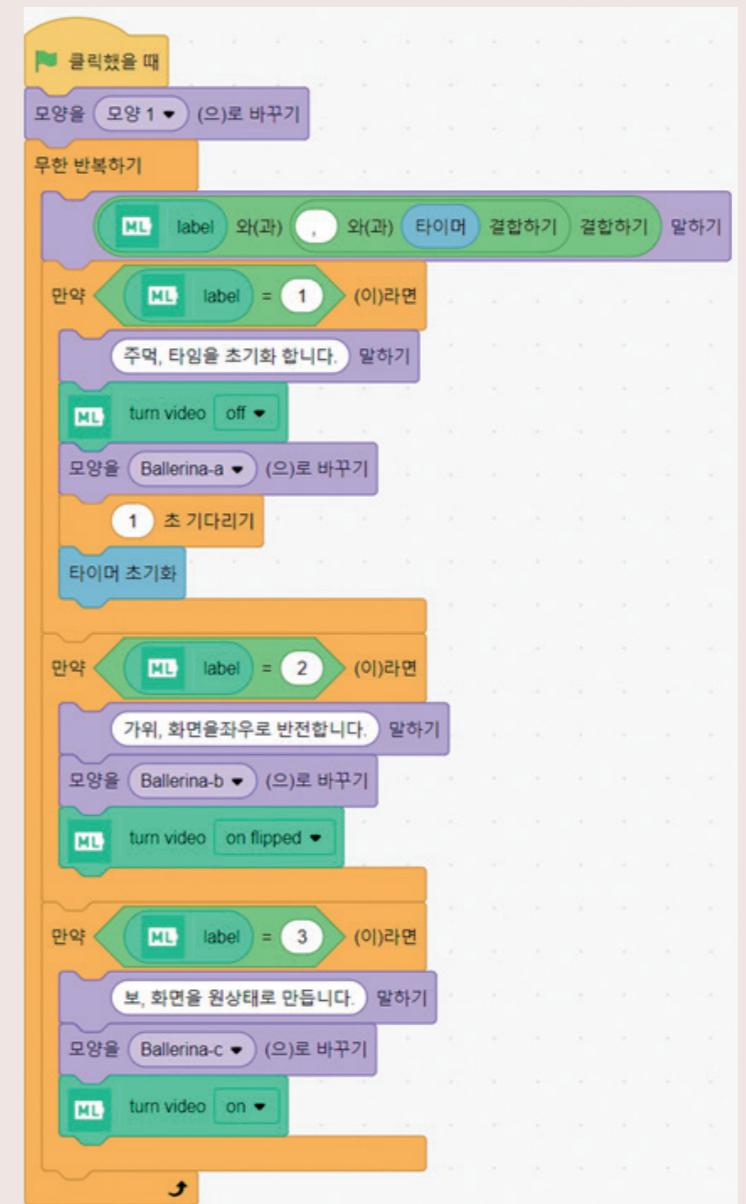
##### ▶ 2단계: 머신러닝 학습시킨 label값 확인하기 (어떻게 할지 적어보세요)



##### ▶ 3단계: 캐릭터 모양 찾기 또는 수정하기 (가위, 바위, 보에 모양과 비슷한 것을 선택)



##### ▶ 4단계: 프로그래밍하기



##### ▶ 5단계: 실행 후 수정하기

수정하고자 하는 내용을 적어보자

# 02

## 핸즈포즈

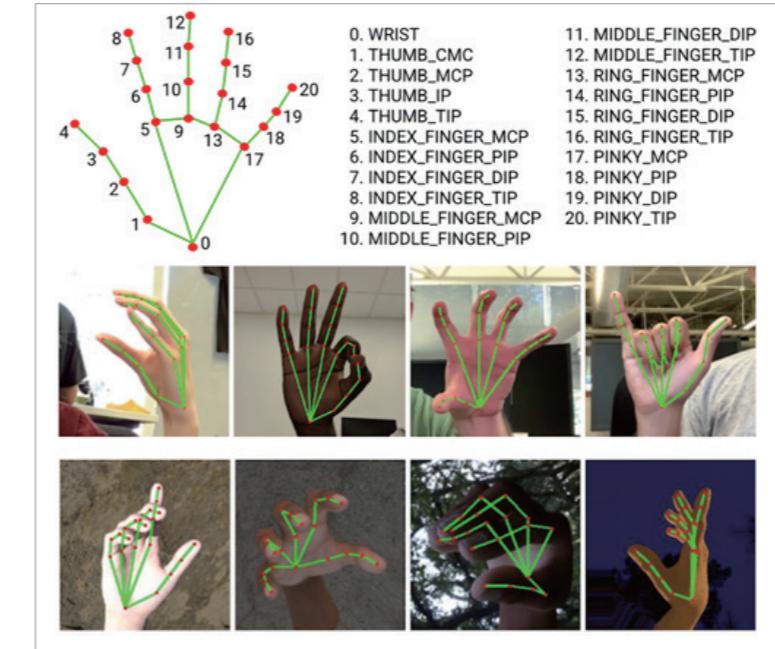
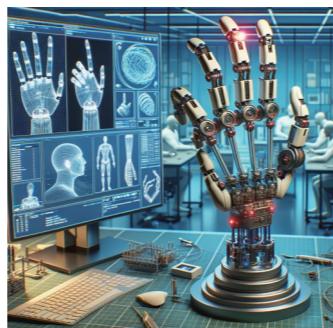


Q. 핸즈포즈로 그림을 그릴 수 있을까?

Q. 인공지능은 어떻게 핸즈포즈를 알수 있을까?

### 1. 핸즈포즈

“핸즈포즈(HandsPose)”는 손의 자세와 동작을 추적하고 인식하는 기술을 일컫는 용어로 사용될 수 있다. 이 기술은 컴퓨터 비전, 인공 지능, 머신 러닝과 같은 분야에서 발전하고 있으며, 특히 손의 3D 포즈 추정에 중점을 둔다. 손가락의 위치, 방향, 그리고 움직임을 실시간으로 파악하여 다양한 응용 분야에 활용한다.



그림\_핸드 랜드마트 모델 활용하기

\* 핸즈포즈 기술의 다양한 방식

단계	설명
입력 데이터 수집	<ul style="list-style-type: none"> <li>카메라나 깊이 센서를 통해 손의 이미지나 비디오 데이터를 수집함.</li> <li>깊이 센서를 사용하면 3D 정보를 얻어 더 정밀한 포즈 추정 가능.</li> </ul>
이미지 전처리	<ul style="list-style-type: none"> <li>손을 이미지에서 분리하고, 이미지를 정규화하는 등의 처리를 통해 데이터를 적절한 형태로 변환함.</li> </ul>
특징 추출	<ul style="list-style-type: none"> <li>알고리즘을 사용하여 손의 중요한 특징(예: 손가락 끝, 관절 위치)을 식별함.</li> </ul>
포즈 추정	<ul style="list-style-type: none"> <li>추출된 특징을 기반으로 머신 러닝 또는 딥러닝 모델을 사용해 손의 3D 포즈를 추정함.</li> </ul>
애플리케이션에 적용	<ul style="list-style-type: none"> <li>추정된 핸즈포즈를 가상현실, 증강현실, 사인 언어 번역, 원격 제어, 의료 재활 훈련, 게임 등 다양한 분야에 적용함.</li> </ul>

## 손가락 명칭

한국어 명칭	영어 명칭
엄지손가락	Thumb
검지손가락	Index Finger
중지손가락	Middle Finger
약지손가락	Ring Finger
새끼손가락	Little Finger, Pinky Finger

## ① 핸즈 포즈(손가락 포즈 인식하기)블록 설정하기

확장에 들어가서 아래의 확장블록을 설정한다.

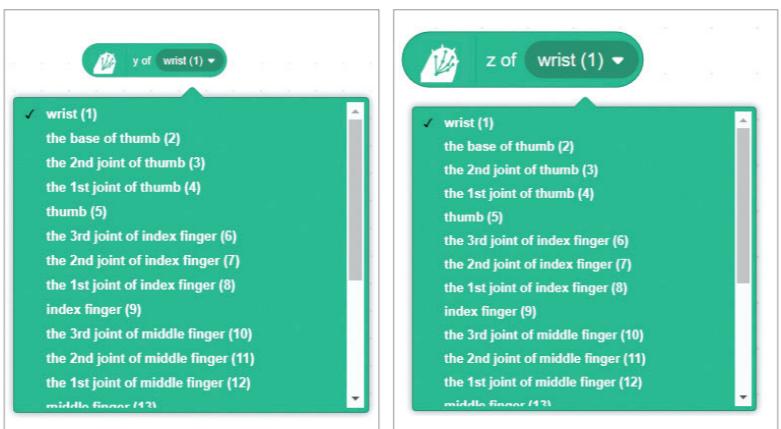
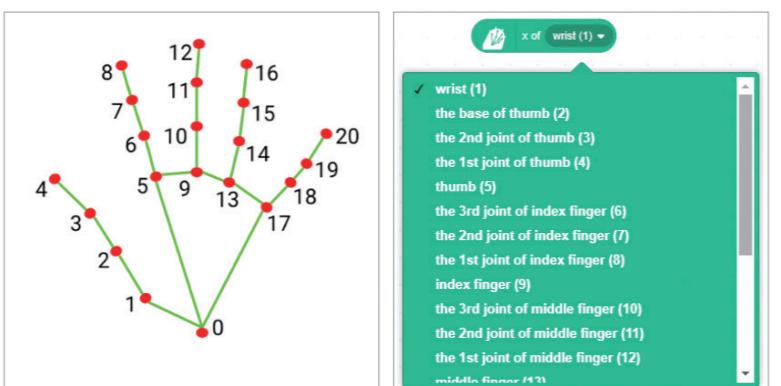


## ② 핸즈 포즈 인식하기 블록 활용하기

핸즈 포즈 인식하기 위해 추가 블록을 활용해 보도록 한다.

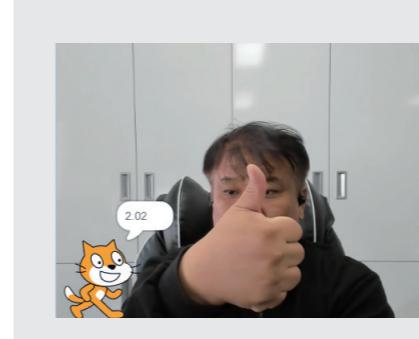
여기에서 중요한 것은 손목 또는 손가락 마디의 위치에 X축, Y축, Z축으로 손가락 마디로 설정하여 표시할 수 있다.

그럼 아래와 같이 손가락 위치를 나타낼 수 있다.

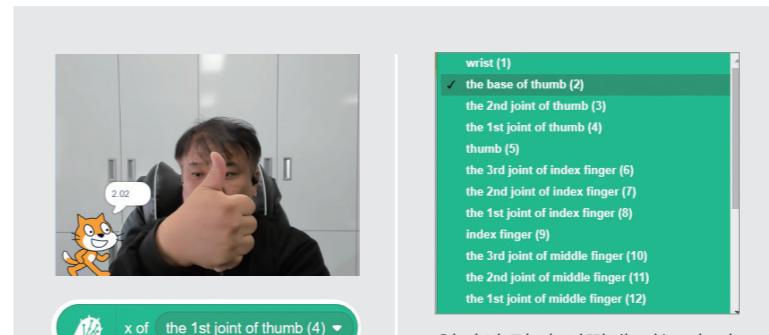


## ③ 손가락 마디 위치 확인하기

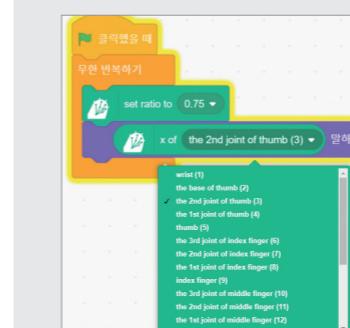
- 핸즈 포즈 동작 중 왼쪽 엄지척 자세를 한다.



- 자세를 취한 후에 각 블록의 값을 변경해 가며 확인해 본다.



콤보박스에 있는 손가락 마디 위치가 나와 있는 설정을 변경해 가면 위치값을 확인해 보자 아래는 손가락 마디 체크 프로그램을 간단하게 작성한다.



위치의 값이 어떻게 되는지 적어보자.

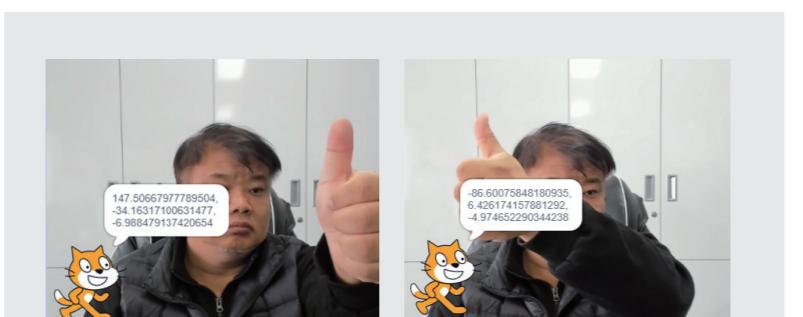


위치의 값이 어떻게 되는지 적어보자.



위치의 값이 어떻게 되는지 적어보자.

- ④ 손가락 마디를 X축, Y축, Z축으로 손가락 마디로 설정하여 표시할 수 있도록 간단하게 프로그램을 아래와 같이 만들어 보자.



엄지척을 한 위치의 값이 변하는 것을 확인해 보세요.

왼쪽의 그림에서의 지점은 X축 147, Y축 -34, Z축 -6의 위치로 되어 있는 것을 알 수 있고 오른쪽그림은 X축 -86, Y축 6, Z축 -4로 변화된 것을 알 수 있다.

이 그림에서 알 수 있듯이 엄지의 마디에 값의 변화가 된다는 것을 알 수 있다.

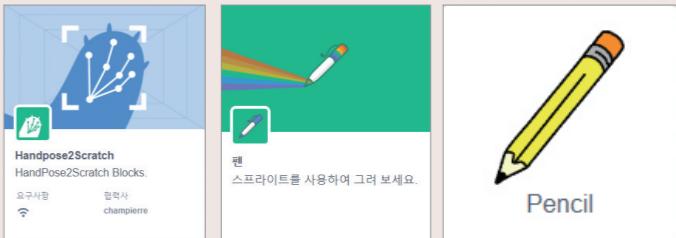

프로젝트 만들기
'손가락으로 그림그리기'

### 1) 기본아이디어

이 프로젝트는 핸드 트래킹 기술, 구체적으로는 '핸즈포즈(HandPose)' 모델을 활용하여 사용자의 검지손가락 위치 데이터를 추적한다. 이 데이터를 통해 가상 환경 내에서 캐릭터가 붓을 사용해 그림을 그릴 수 있도록 구현된다. 사용자의 검지손가락 Z축 위치에 따라, 붓이 두 가지 다른 색상을 사용하여 그림을 그릴 수 있도록 할 수 있을 것이다. 이 기능을 실현하기 위한 단계별 접근 방식을 통해 프로젝트를 개발하고자 한다.

#### ▶ 1단계

개발하기 위한 확장 블록, 그리기 블록 그리고 스프라이트 아이콘 만들기를 설치한다.



#### ▶ 2단계

프로젝트에 필요한 블록을 준비한다.



검지손가락 X, Y, Z축



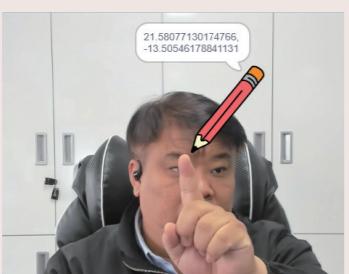
펜의 색상을  
빨간색, 노란색,  
굵기, 지우기 블록



스프라이트 펜의 색상 변경하기 위해 그림작업을 합니다. 여기에서 각 객체의 그림은 펜의 연필심의 위치를 중앙에 위치

## ▶ 3단계

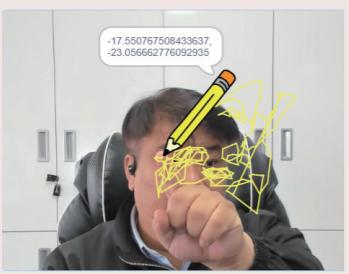
기본 프로그래밍 하기 위해 검지손가락 위치의 값을 알 수 있도록 블록을 코딩한다.



- 블록을 활용하여 검지손가락에 따라 다니는 펜을 만들기 위해 이동블록을 사용

## ▶ 4단계

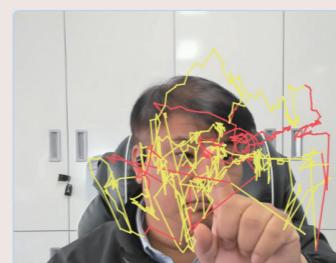
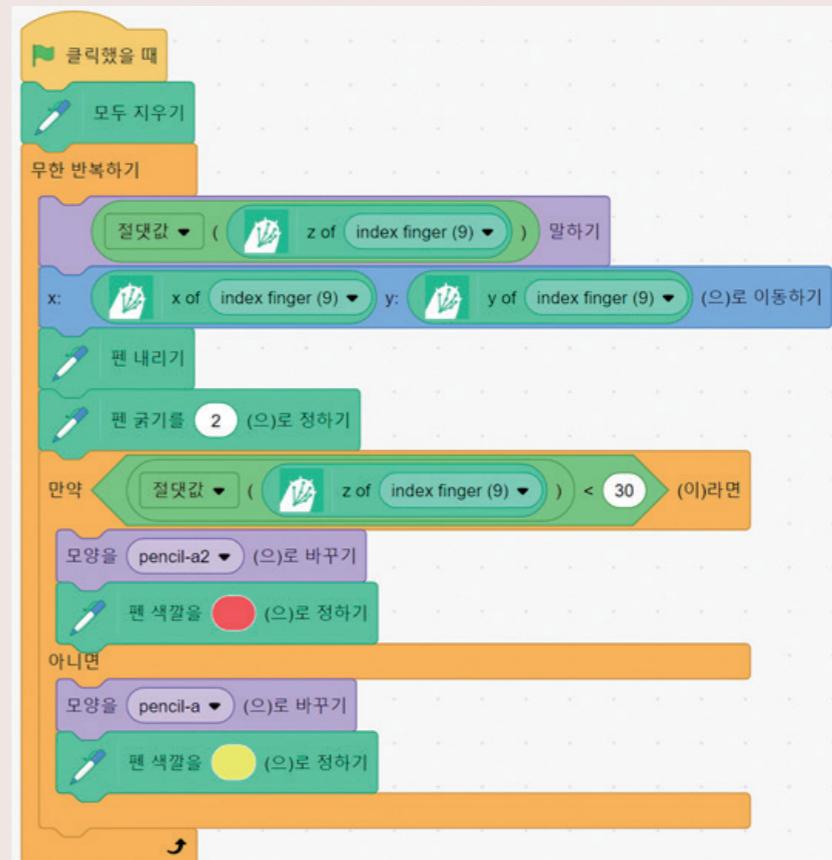
검지손가락으로 펜으로 그림 그리기 위해 아래와 같이 만들어보자



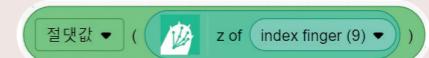
- 펜의 색은 노란색으로 굵기는 2로 설정

## ▶ 5단계

검지손가락을 Z축 위치값을 활용하여 색상과 캐릭터의 모양을 변경하도록 만들어보자.



- Z축 값의 위치값을



- 블록의 위치값이 30미만일 때 빨간색 펜이 빨간색을 그리도록 하고 기본적으로 노란색 펜으로 그림을 그리기

## ▶ 6단계 : 실행 후 수정하기

수정하고자 하는 내용을 적어보자

### III. AI로 디바이스 제어

이 단원에서는 노코드 AI영상처리 서비스를 이용해 웹페이지의 오브젝트를 제어해 보고, 시리얼통신으로 연결된 아두이노를 직접 제어해 보는 실습을 통해 AI로 모니터안의 세상 뿐만 아니라 모니터 밖의 세상까지 제어 가능함을 체험해 보면서 AI가 온/오프라인 어디나 영향을 미친다는 것을 알 수 있는 경험을 갖게 한다.

1. 티쳐블머신 웹페이지 제어
2. 티쳐블머신 아두이노 제어

01

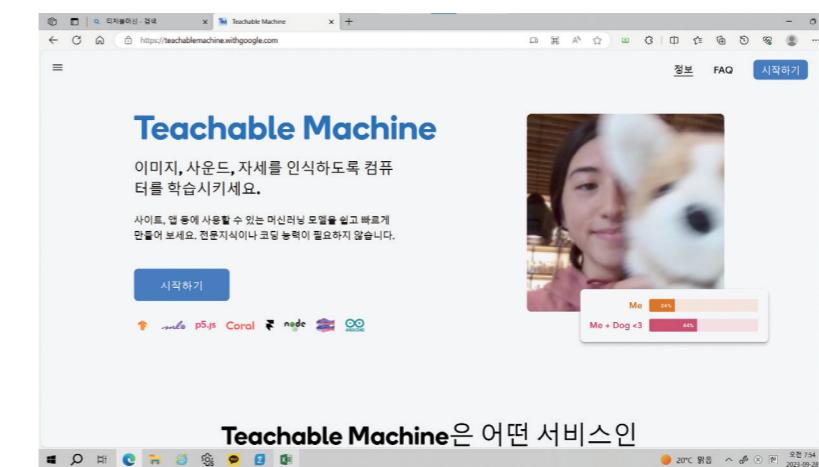
#### 티쳐블머신 웹페이지 제어

: 티쳐블 머신의 이미지, 음성, 포즈 기능을 이용한 초간단 웹 AI 애플리케이션 만들기

##### 1. 티쳐블 머신(Teacherble Machine)

티쳐블 머신은 머신러닝과 코딩에 대한 지식 없이도 머신러닝을 이용할 수 있게 고안된 도구이다. 티쳐블 머신을 이용하여 AI로 도형을 분류하는 모델을 만들어보자. 검색창에 '티쳐블 머신'을 검색하거나 아래 주소를 직접 입력한다. (크롬 브라우저 권장)

- 사이트 주소: <https://teachablemachine.withgoogle.com/>



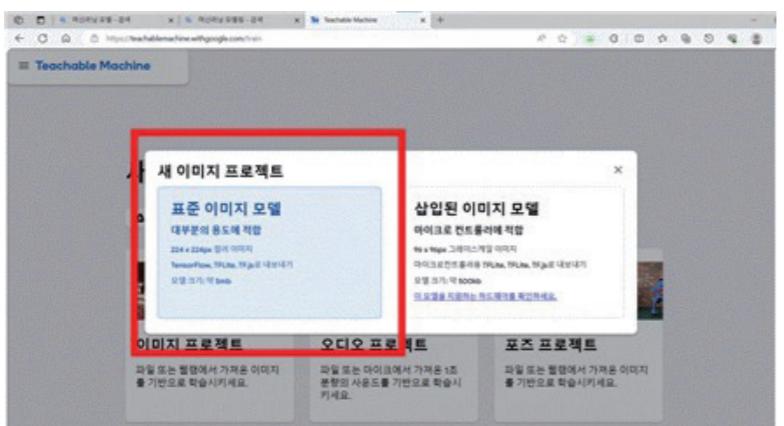
티쳐블 머신 첫 화면 구성

머신 러닝 모델이란 데이터 세트에서 패턴을 찾거나 이를 근거로 결정을 내릴 수 있는 프로그램으로 이 모델을 만드는 과정을 학습이라고 한다.

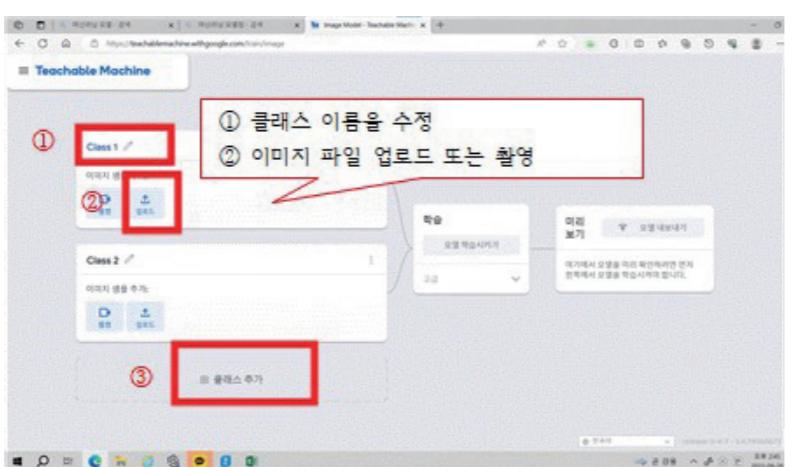
'시작하기'를 클릭하면 이미지, 오디오, 포즈를 인식하도록 컴퓨터를 학습시키고 모델링할 수 있다. 이미지와 포즈 프로젝트를 이용하면 이미지 파일 또는 웹캠이 필요하며, 오디오 프로젝트는 소리파일 또는 마이크가 필요하다.



▶ 이미지 프로젝트를 클릭한 후 표준 이미지 모델을 선택한다.

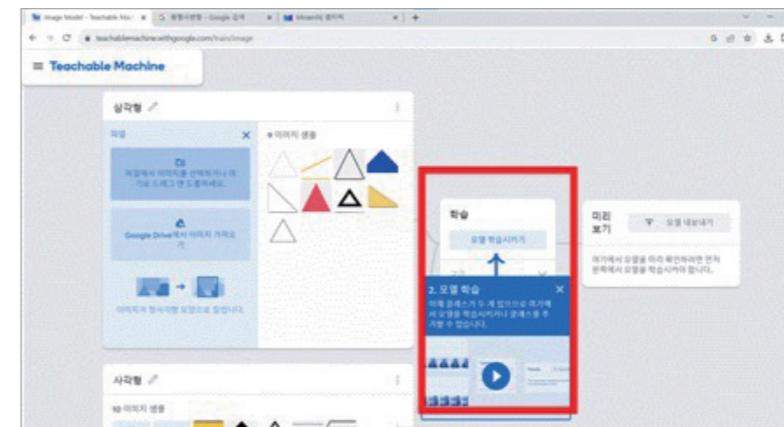


아래 화면에서 이미지를 학습시키고 모델을 만들 수 있다.

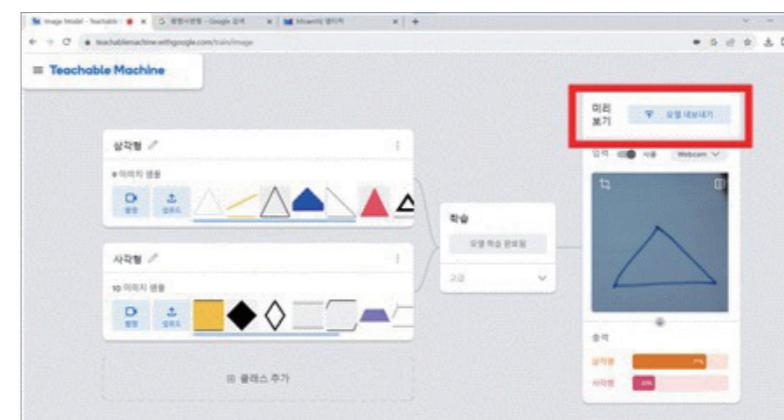


클래스(class)는 같은 종류의 이미지들을 모아서 그룹화하여 이름을 붙인 것이다. 간단하게 도형을 분류하는 모델을 만들어 보자.

- ▶ class 1이라고 적힌 부분에 '삼각형'이라고 입력하고 class 2에는 '사각형'을 입력한다. 3개 이상의 클래스를 만들고 싶으면 '클래스 추가'를 클릭한다.
- ▶ 업로드를 클릭한다. 다음과 같은 화면에서 PC나 드라이브에 저장된 삼각형, 사각형 그림 파일을 올린다. 정확한 학습을 위해서는 사진이 많을수록 좋다.



- ▶ 각각의 클래스에 사진 파일 업로드가 되었으면 모델 학습시키기를 클릭한다. 이 단계에서 컴퓨터는 각각의 클래스에 속한 여려 사진을 보고 특징을 학습한 후 분류할 수 있게 된다.
- ▶ 학습이 끝나면 웹캠의 권한을 '허용'하고 미리보기 창에서 결과를 확인할 수 있다.

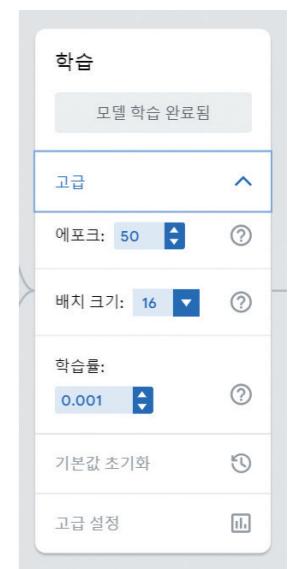


이 단계에서 중요한 것은 기계가 제대로 학습을 했는지 평가하는 것이다. 평가 결과가 부정확다면 이미지를 추가한 후 다시 학습시킨다.

도형 그림 파일 다운로드 링크:  
<https://bit.ly/tm-device>  
또는 검색 엔진을 이용해 삼각형과 사각형 그림 파일을 찾아 각각 10장 이상 PC에 저장한다.

업로드가 아닌 웹캠을 클릭하면 카메라 사용 권한을 '허용'하고 길게 눌러서 녹화하기를 계속 클릭하면 여러 장면을 연속해서 촬영할 수 있다.

모델 학습시키기 아래의 고급 탭을 클릭하면 에포크, 배치 크기, 학습률 등을 설정할 수 있다.

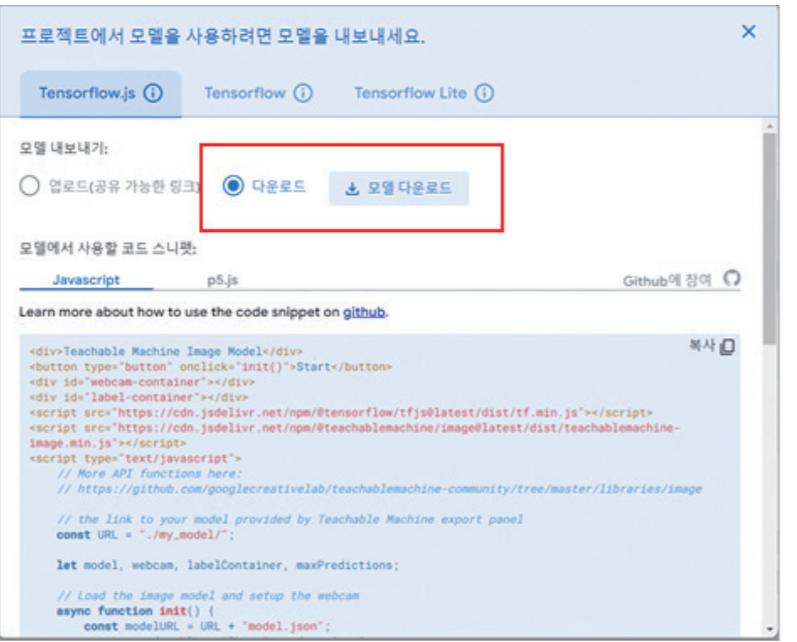


삼각형 그림을 웹캠에 보여주면 머신 러닝이 판단한 결과도 삼각형으로 출력된다.

이런 학습 과정을 거쳐서 컴퓨터는 정교한 판단력을 갖게 되고 이런 기술을 기계학습 또는 Machine Learning이라고 부른다.

다운로드가 아닌 업로드를 이용하면 모델의 공유 링크가 생성되고 이 링크를 복사해서도 공유 및 코딩을 할 수 있다.

- ▶ 미리보기 창의 모델 내보내기를 클릭한다.
- ▶ 다음 그림과 같이 팝업 창이 나타나면 Tensorflow.js 탭에서 다운로드를 선택하고 모델 다운로드를 클릭한다.



압축 풀기를 해야 나중에 이 파일들을 사용할 수 있다.

#### json

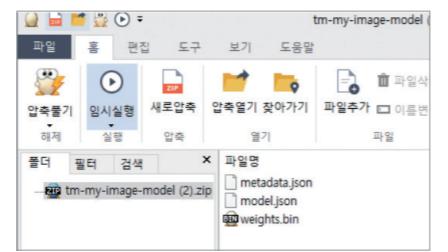
키-값 쌍으로 이루어진 데이터 오브젝트를 전달하기 위해 인간이 읽을 수 있는 텍스트를 사용하는 개방형 표준 포맷

#### bin

바이너리 형식의 데이터 프로그램을 격납하고 있는 파일

- ▶ tm-my-image-model.zip 이름의 압축 파일이 다운로드 되고 압축 풀기를 실행한다.

압축을 풀면 폴더 안에 metadata.json, model.json, weights.bin 세 개의 파일이 있다. 이 파일을 이용해서 간단한 웹 애플리케이션을 만들어보자.

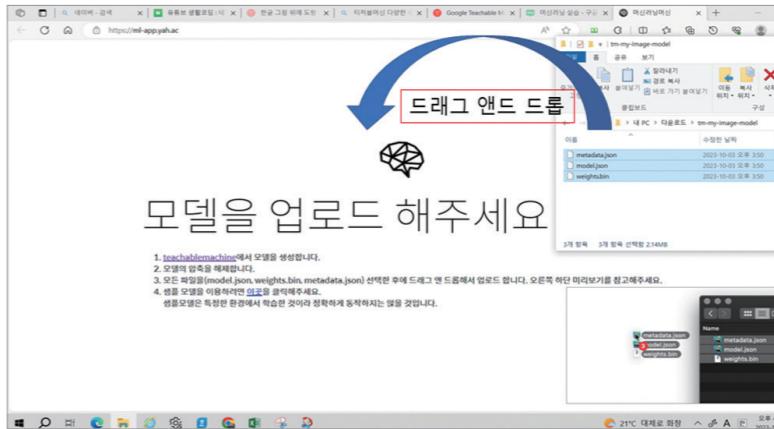


## 2. 초간단 마신러닝 웹 애플리케이션 만들기

티처블 머신으로 만든 도형 분류 모델을 이용하여 인식 결과를 메시지로 알려주고 음성으로 말하는 앱을 만들 수 있다. 유튜브 '생활코딩'에서 소개한 머신러닝 머신 서비스를 이용을 위해 아래 주소로 방문한다.

- 사이트 주소: <https://ml-app.yah.ac/> (크롬 브라우저 권장)

- ▶ 티처블 머신에서 다운로드 했던 tm-my-image-model.zip의 압축을 폰 폴더에 있는 3개의 파일(metadata.json, model.json, weights.bin)을 모두 선택한 후 화면으로 드래그 앤드 드롭을 한다.
- ▶ 파일들이 모두 업로드되면 카메라 사용을 '허용'한다.



업로드가 완료되면 블록 코딩으로 티처블 머신의 인식 결과를 화면에 보여주고 음성으로 말하는 웹 애플리케이션을 만들어 보자.



- ▶ 인식 결과가 70%보다 크면 메시지로 표시하고 음성으로 말하도록 위와 같이 블록을 드래그하고 회색 영역을 수정하여 코딩을 완료한다.

유튜브 생활 코딩 채널  
<http://opentutorials.org/course/1>

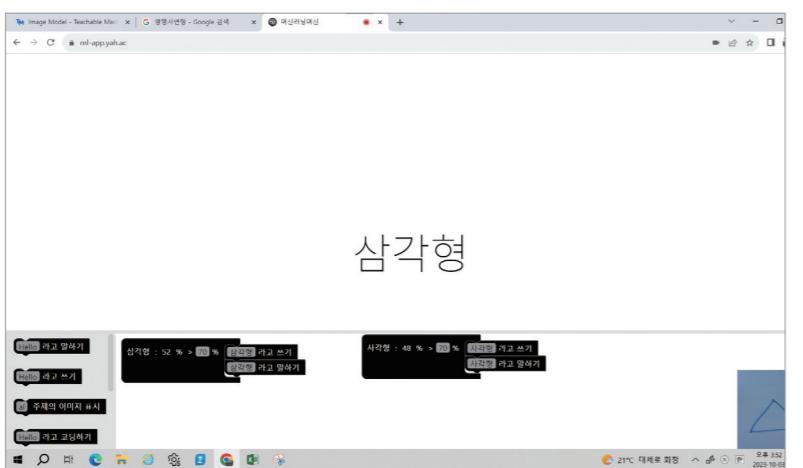
왼쪽의 퍼센트는 웹캠에 찍히 화면이 클래스에 속할 확률을 나타내며 오른쪽 빙칸에 넣은 숫자보다 확률이 높으면 화면에 끼운 블록이 실행된다.



02

## 리처블머신 아두이노 제어

: 리처블 머신의 학습 모델을 이용해 다양하게 작동하는 아두이노 장치 만들기.



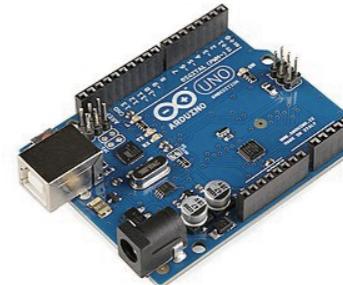
이제 삼각형과 사각형을 분류하여 알려주는 초 간단 웹 애플리케이션이 완성되었다.

이와 같은 방법으로 이미지, 음성, 포즈 프로젝트를 활용한 AI 분류 모델을 티처블 머신으로 만들고 간단한 애플리케이션을 제작해 보자. 수학에서는 도형과 상관관계 분류, 체육에서는 올바른 운동 자세, 역사에서는 문화 유산 구분하기, 과학에서는 암석 분류하기, 음악과 악기 소리 분류 등으로 활용할 수 있다.

### 1. 아두이노(Arduino)

아두이노는 대표적인 마이크로 컨트롤러(MCU)의 종류로 다양한 장치와 연결하여 입출력할 수 있으며 작고 저렴하지만 다양하게 활용될 수 있다는 장점이 있다.

고릴라 실드를 사용하면 아두이노와 각종 부품을 쉽게 연결하여 사용할 수 있으므로 추천한다.



아두이노만 있을 때

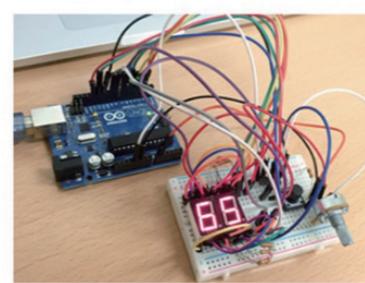
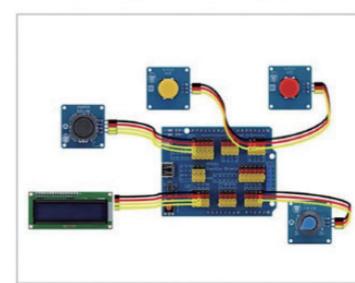


사진 출처\_도서 '시간순삭 인공지능 with 엠블록 & 아두이노'

아두이노에 고릴라 실드를 연결했을 때



고릴라 실드 구입처:  
<https://www.gorillacell.kr/>

AIoT 기술이 적용된 전동 킥보드 유튜브 영상:  
<https://youtu.be/KFzpxA29epo?feature=shared>

## 2. AIoT

AIoT는 'Artificial Intelligence of Things', 즉 인공지능(Artificial Intelligence)과 사물인터넷(IoT : Internet of Things)의 합성어로, 한국어로는 '사물지능' 또는 '사물지능융합기술'이다. 즉, AIoT는 사물인터넷 디바이스에 AI 기법이 적용된 것이다. IoT와 AIoT는 모두 사물에서 센싱한 데이터의 수집과 데이터 처리 과정에서는 동일한 기능을 수행한다. 다만, IoT는 정의된 프로그램과 알고리즘 기반으로 동작하지만, AIoT는 학습과 추론에 의한 지능형 동작을 수행한다는 것이 차이점이다. 대표적인 사례로는 자율주행과 스마트 팜 등이 있다.



클래스 이름을 단순한 숫자로 설정하면 코딩하기가 수월하다.

클래스별 포즈의 수를 일정하게 하는 것이 좋다.

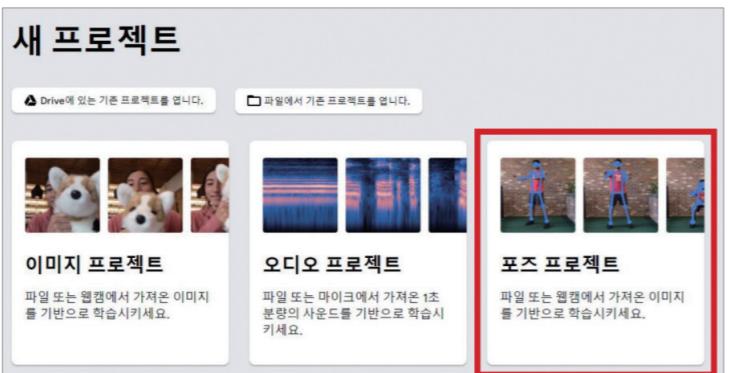
## 3. AIoT 제작하기

앞에서 배웠던 티처블 머신으로 AI 모델을 만들고 포즈에 따라 움직이는 AIoT를 만들어보자. 준비물은 아두이노, 고릴라 실드, LCD디스플레이, 네오픽셀, 웹캠이 필요하다.

### ① 티처블 머신으로 AI 모델 만들기

티처블 머신에 대한 자세한 설명과 사용법은 앞 챕터에도 있다.

- ▶ 티처블 머신(<https://teachablemachine.withgoogle.com/>)에서 포즈 프로젝트를 선택한다.

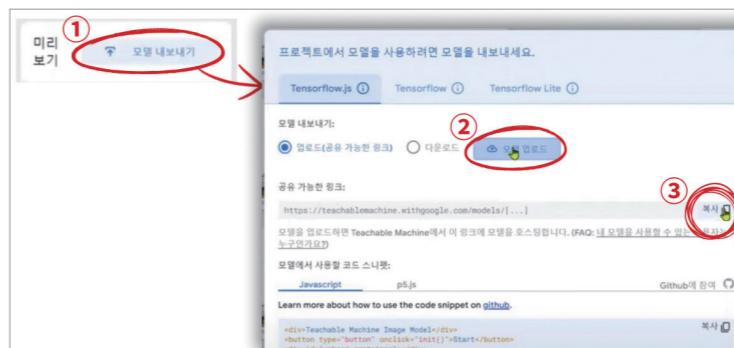


- ▶ 웹캠을 클릭하고 길게 눌러서 녹화하기로 클래스별로 자세를 다르게 하여 저장한다. 아래와 같이 클래스 '1'은 정자세, 클래스 '2'는 왼팔 들기, 클래스 '3'은 팔을 오른팔을 들어 녹화한다. 조금씩 다른 자세로 여러 장 촬영한다.

- ▶ 각각의 클래스에 포즈 파일 업로드가 되었으면 모델 학습시키기기를 클릭한다. 학습이 끝나면 웹캠의 권한을 '허용'하고 미리보기 창에서 결과를 확인할 수 있다. 학습이 잘 되었는지 확인한다.



- ▶ 미리보기 창의 모델 내보내기를 클릭한다. 팝업 창이 나타나면 업로드를 선택하고 모델 업로드를 클릭한다. 공유 가능한 링크가 생성되면 주소를 복사한다.



웹캠을 사용 중지해야 나중에 아두이노와 통신하는 페이지에서 티처블 머신 모델을 불러와 카메라로 작동할 수 있다.

- ▶ 아두이노와 연결을 위해서 링크 복사 후 웹캠을 꼭 사용 중지 시킨다.

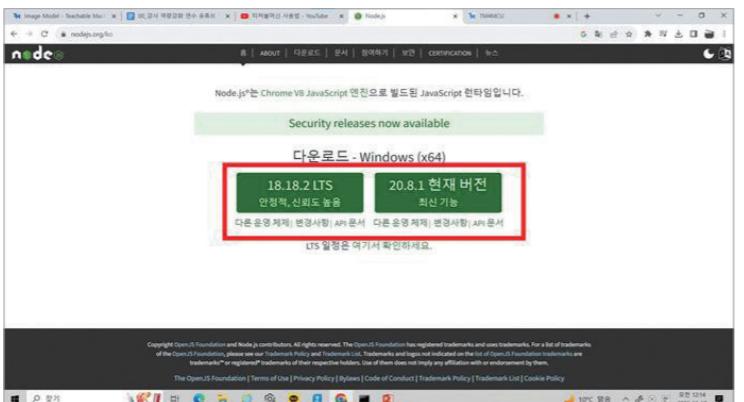


## ② 시리얼 통신을 위한 웹페이지 구현하기

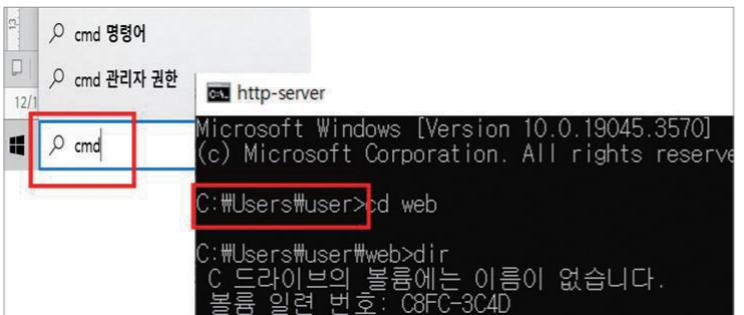
앞에서 만든 티처블 머신 모델과 디바이스의 통신을 위해 가상의 웹서버를 만들고 티처블 머신의 모델에 맞는 HTML파일을 불러와야 한다.

가상의 서버를 구축하면 여러 대의 카메라를 사용할 때 유용하다.

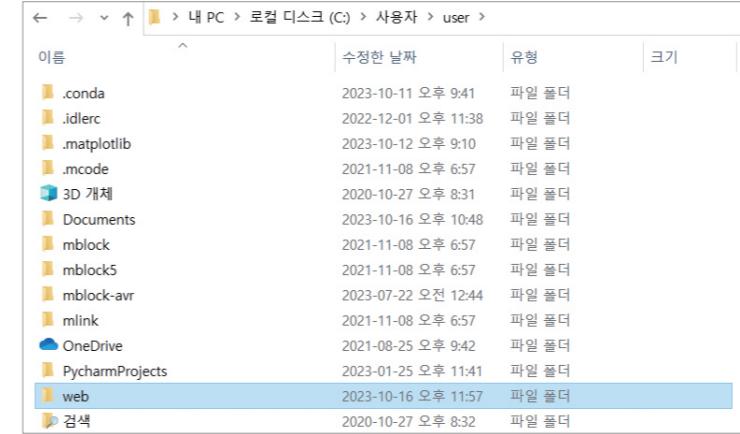
- ▶ 웹서버 구축을 위해 검색창에 node.js를 검색하거나 <https://nodejs.org/ko> 접속한다. 두 가지 버전 중 한 가지를 선택해 내려받고 설치한다.



- ▶ 명령 프롬프트를 실행한다. 작업표시줄의 찾기에서 'cmd'를 입력하면 된다.



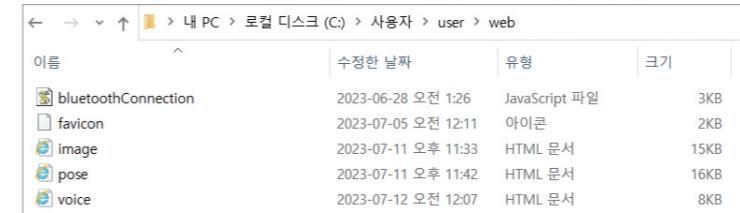
- ▶ 위와 같이 명령 프롬프트 창이 뜨면 파일 탐색기에서 첫 줄의 경로가 있는 폴더로 이동한다. 해당 폴더 안에 새 폴더 만들기 후 폴더명을 'web'으로 변경한다.



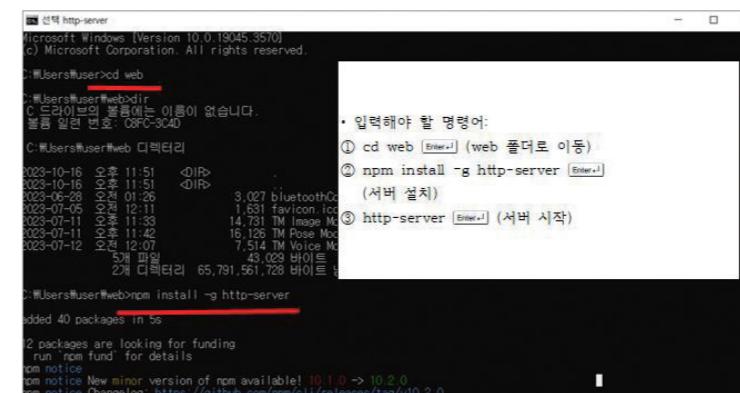
첫 줄의 경로는 사용자마다 다르다.

마우스 오른쪽 클릭-새로 만들기-폴더를 클릭하여 새폴더를 만든다.

- ▶ web폴더 안에 TM 3Models4MCU 폴더 안의 파일들을 복사-붙여 넣기 한다. 붙여넣기 후 웹페이지 호출을 쉽게 하려면 각각의 파일명을 'image.html','pose.html','voice.html'과 같은 간단한 이름으로 바꾼다.



- ▶ 다시 명령 프롬프트로 돌아가서 가상의 웹서버를 설치한다. 아래처럼 순서대로 명령어를 입력한다.



TM 3Models4MCU 폴더 파일 다운로드 링크:

<https://bit.ly/tm-device>

Ctrl+C를 치면 서버가 정지된다.

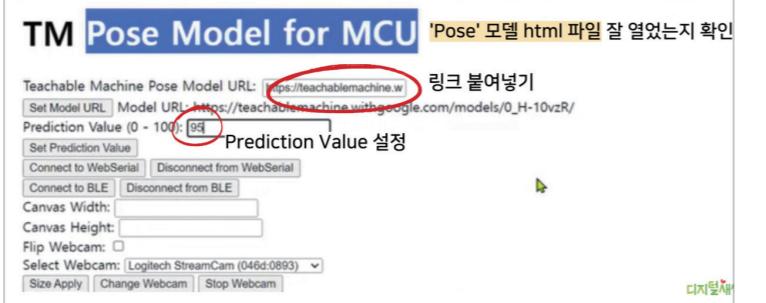
앞에서 티쳐블 머신 포즈 프로젝트를 만들었으므로 pose.html을 불러온다. 각 프로젝트에 맞는 페이지를 불러오면 된다.

```
선택 http-server
! /Users/user/web/htp-server
starting up http://localhost:8080 ...
http-server version: 14.1.1
no server settings:
  OPS: disabled
  cache: 3600 seconds
  connection Timeout: 120 seconds
  directory Listings: visible
  gzip: enabled
  serve GZIP Files: false
  serve Brotli Files: false
  default File Extension: none
available on:
  http://192.168.45.84:8080
  http://127.0.0.1:8080
hit CTRL-C to stop the server
2023-10-16T14:59:18Z [GET /image/] "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/118.0.0.0 Safari/537.36" [DEPRECATION] DeprecationWarning: OutgoingMessage.prototype._headers is deprecated
2023-10-16T14:59:18Z [GET /image/] "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/118.0.0.0 Safari/537.36" Error (404): "Not found"
2023-10-16T14:59:18Z [GET /image/] "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/118.0.0.0 Safari/537.36" [DEPRECATION] DeprecationWarning: OutgoingMessage.prototype._headers is deprecated
2023-10-16T14:59:18Z [GET /image/] "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/118.0.0.0 Safari/537.36"
```

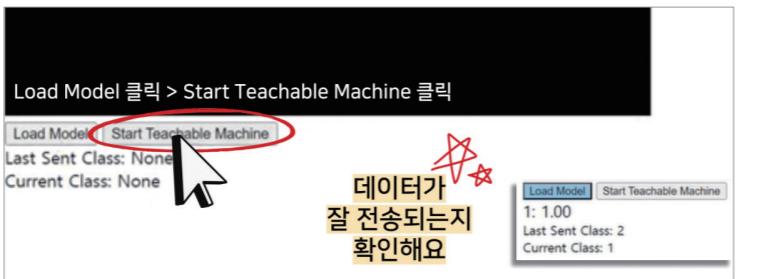
- ▶ 웹 브라우저 주소 입력창에 'http://localhost:8080/pose'를 입력하면 티쳐블 머신과 아두이노를 통신할 수 있는 TM Pose Model for MCU 웹 페이지가 보인다.

Canvas Width와 Height로 카메라 창의 사이즈를 변경하고 Size Apply로 적용한다.

• 여러 대의 카메라를 사용할 경우 Change Webcam으로 선택할 수 있다.



- ▶ 위와 같은 웹페이지가 보이면 앞에서 복사했던 티쳐블 머신의 모델 링크를 붙여넣기 후 Set Model URL을 클릭한다. Prediction Value는 약 95 정도로 한다.



- ▶ 카메라 창 아래에 있는 Load Model을 클릭한 후 Start Teachable Machine을 클릭한다. 카메라를 허용하고 포즈를 취하여 클래스가 잘 분류되는지 확인한다.

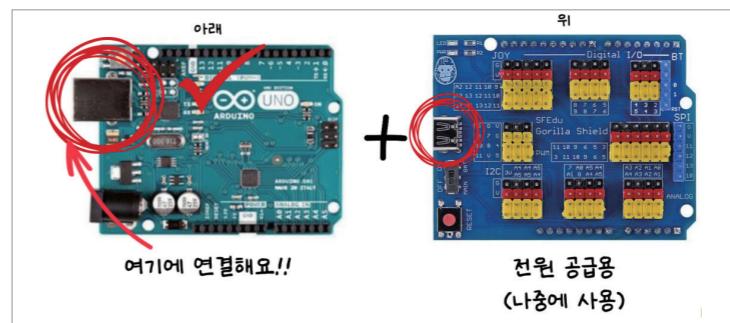


※ 창을 닫지 마세요.  
아두이노와 연결을 해야 합니다.

## ② 아두이노 장치 준비하기

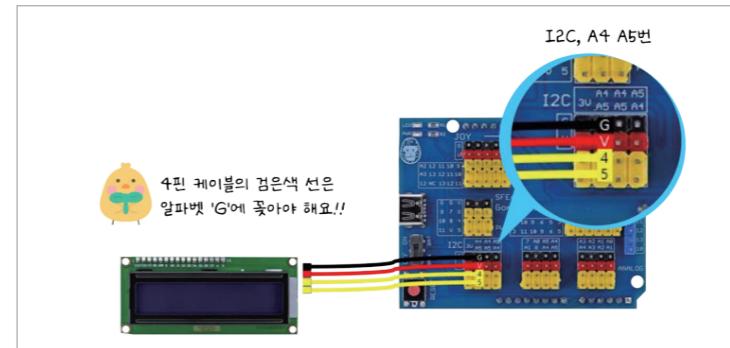
티쳐블 머신으로 AI모델을 완성했으면 이제 디바이스를 만들어보자.

- ▶ 아두이노 위에 고릴라 실드를 결합한 후 C타입 케이블을 이용하여 컴퓨터 USB와 연결한다.



아두이노와 고릴라 실드를 핀과 흠을 잘 맞추어 결합한다.

- ▶ 아래와 같이 I2C LCD 디스플레이 장치를 점퍼선으로 연결한다. 티쳐블 머신에서 인식한 결과 값을 출력할 수 있다.

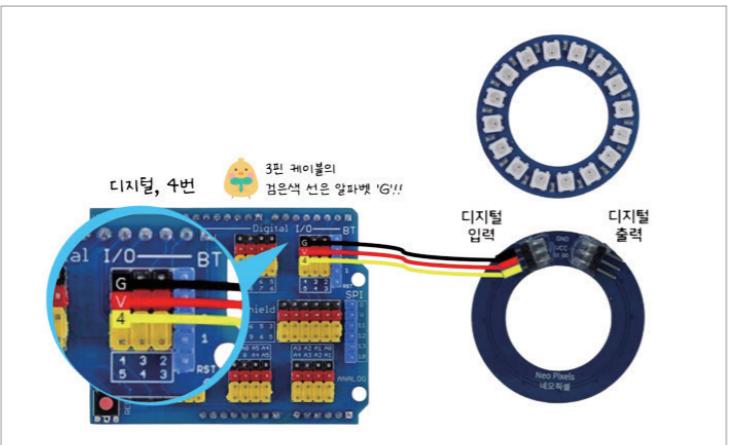


일반적으로 G(GND, Ground)는 검은 선으로 연결하고 V(Vcc)는 빨간색 선으로 연결한다.

LCD 디스플레이 뒷면에 점퍼선과 연결하는 핀이 4개 있다.

- ▶ 16개의 LED에 빛을 출력하는 네오픽셀은 디지털 신호 장치이므로 고릴라 실드에서 Digital I/O 부분에 그림과 같이 연결한다.

네오픽셀 뒷면에 점퍼선을 꽂는다.

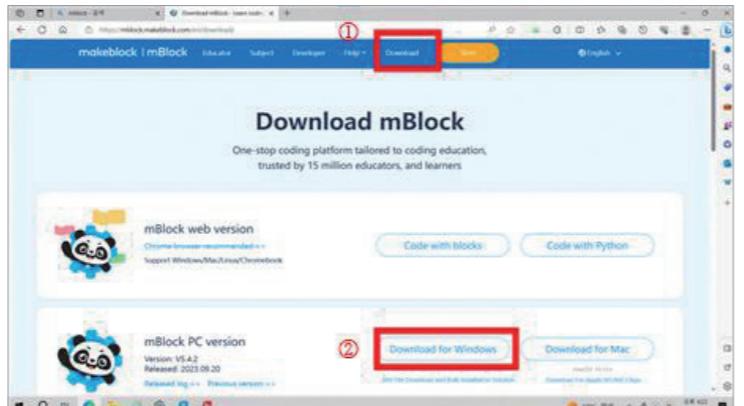


### ③ 블록 코딩하기

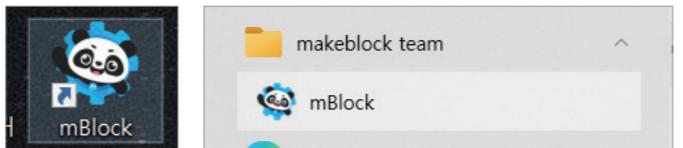
장치 연결이 완료되면 mBlock 사이트로 이동한다. mBlock은 인공지능과 사용자 확장 블럭을 사용할 수 있어 매우 유용하다.

- 사이트 주소: <https://mblock.makeblock.com/en/>

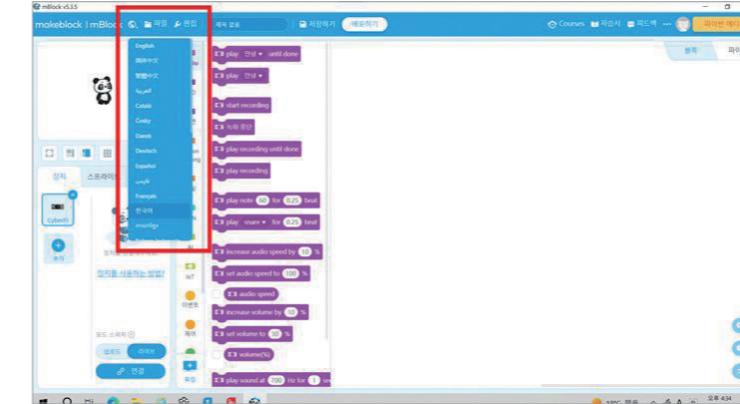
- ▶ 홈페이지 상단 메뉴에서 다운로드를 선택 후 PC 운영체제 사양에 맞는 버전을 다운로드 한다.



- ▶ 프로그램 설치를 완료하면 바탕화면의 바로가기 아이콘이나 프로그램 메뉴에서 mBlock을 선택한다.



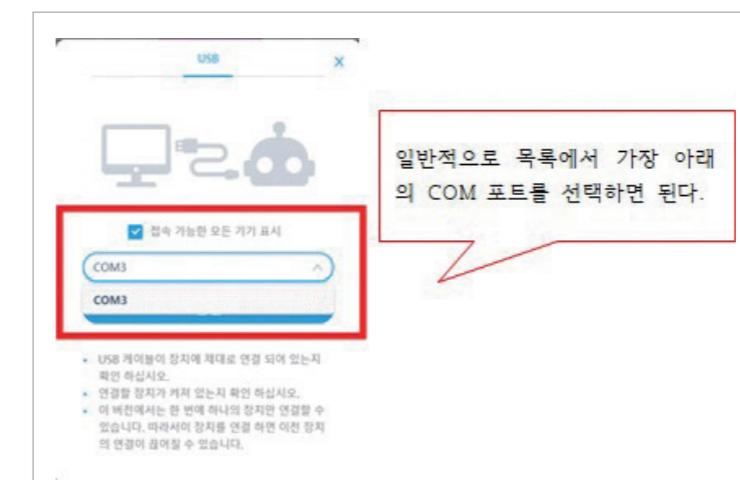
- ▶ 아래와 같은 화면에서 지구 모양의 메뉴를 클릭하고 한국어로 변경 한다.



- ▶ 왼쪽 하단의 장치 탭에서 기본적으로 연결된 장치인 CyberPi는 제거하고 아래와 같이 아두이노 장치를 추가한 후 연결을 선택한다.



- ▶ 업로드 모드를 선택한 후 연결을 클릭하면 팝업이 뜬다. 접속 가능한 모든 기기를 표시하고 아두이노가 연결된 포트 번호를 선택한다.



오른쪽 모서리 위의 ⊗를 클릭하여 장치를 제거한다.



업로드 모드: 장치에 코드를 업로드하여 장치가 PC와 연결이 해제된 이후에도 프로그램 코드에 따라 작동됨.

라이브 모드: 실시간으로 코드를 디버깅하여 실행 여부를 확인할 수 있음.

고릴라 실드와 여러 센서와 입출력 장치 사용을 위해 확장 블록을 설치해야 한다.

- ▶ 다음과 같이 블록 구성 메뉴에서 확장을 클릭하고 '공공두이노'와 '유ти리티셀'을 검색한다.

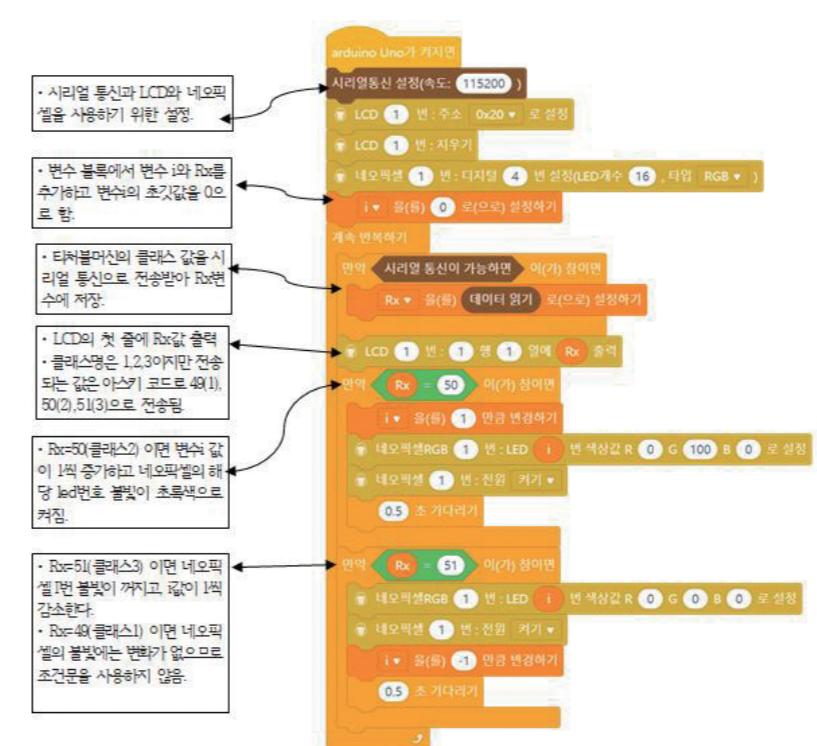


- ▶ 공공두이노와 유ти리티 셀을 추가 버튼을 클릭하여 설치한다.

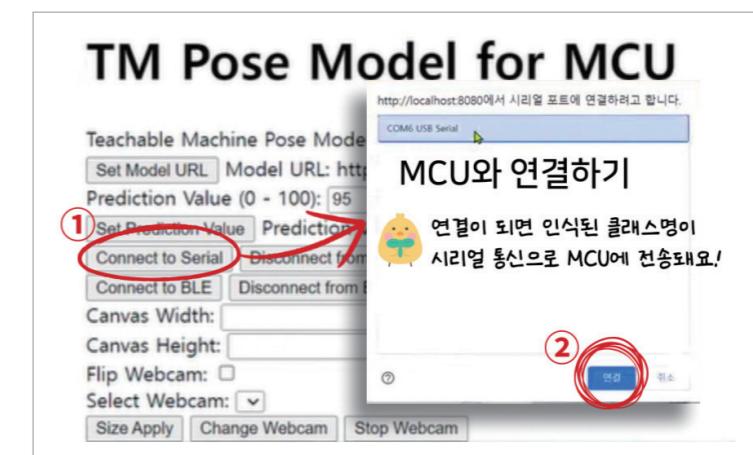


mBlock 코딩 파일 다운로드 링크:  
<https://bit.ly/tm-device>

- ▶ 티처블 머신의 인식한 클래스 값을 LCD 디스플레이에 출력하고 인식한 '클래스 1'(Rx=49)이면 네오픽셀 불빛에 변동이 없고 '클래스 2'(Rx=50)이면 빛이 증가하고 '클래스 3'(Rx=51)이면 빛이 꺼지는 프로그램을 아래와 같이 코딩한다.



- ▶ 코딩이 완료되면 업로드 후 연결 해제를 클릭한다. 앞에서 티처블 머신과 연동한 TM Pose Model for MCU 웹 페이지로 돌아가 Connect to Serial을 클릭하고 아두이노가 연결된 포트를 선택하여 연결한다.

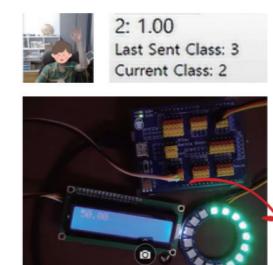


mBlock 코딩 파일 다운로드 링크:  
<https://bit.ly/tm-device>



연결 해제를 해야 티처블 머신과 연동하는 웹페이지에서 장치와 연결을 할 수 있다.

카메라로 인식된 클래스 값이 시리얼 통신으로 아두이노에게 전송되고, 전송된 데이터가 LCD에 출력되면 값에 따라 네오픽셀 불빛이 커지고 꺼지게 할 수 있다. 티처블 머신의 포즈 모델에 의해 장치가 잘 작동되면 성공이다.





티처블 머신과 아두이노를 이용하여 다음과 같이 다양한 프로젝트를 할 수 있다.

- 재활용 쓰레기를 종류별로 인식하여 분리해서 수거하는 장치
- 공부할 때 즐거나 잠들면 깨워주는 장치
- 악기 소리에 따라 분류하여 작동하는 장치
- 물건의 색상에 따라 구분하여 분류하는 장치

## '모든 교사를 위한 인천형 인공지능' 연수 동영상

2022 개정교육과정에 대비하여 디지털교육 역량강화에 대한 강조와 더불어 인공지능에 대한 개념과 원리부터 적용까지 아우르는 연수가 요구되고 있다. 본 영상은 '모든 교사를 위한 인천형 인공지능' 활용 라이브 연수 녹화영상으로 교재를 활용하고자 하는 교사들이 주요 개념 및 활용법을 보다 쉽게 이해할 수 있도록 교재에 싣게 되었다.



2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 6일...  
2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 5일...  
2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 3일...  
2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 2일...  
2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 1일...  
2023 모든 교사를 위한 인천형 인공지능 교재 활용 연수 1일...

교회수 29회  
• 스포리얼 시간: 1개월 전

교회수 182회  
• 스포리얼 시간: 1개월 전

교회수 104회  
• 스포리얼 시간: 1개월 전

교회수 20회  
• 스포리얼 시간: 1개월 전

교회수 417회  
• 스포리얼 시간: 2개월 전

교회수 417회  
• 스포리얼 시간: 2개월 전

**1일차**  
1교시 인공지능이란 무엇인가?  
2교시 ChatGPT 활용하기 1 (텍스트 활용, 뷰튼-스튜디오&스토어)  
<https://www.youtube.com/watch?v=6bcRKfLkpto>



1일차

**2일차**  
1교시 인공지능 테크놀로지가 뭐지?  
2교시 응용프로그램을 만들어보자(ml2 스크래치와 핸즈포즈)  
<https://www.youtube.com/watch?v=-qRo6o-Lhlw>



2일차

**3일차**  
1교시 데이터, 그것을 알고 싶다  
2교시 ChatGPT 활용하기 2 (문서요약, 이미지 생성하기)  
<https://www.youtube.com/watch?v=yDt67t13pgs>



3일차

**4일차**  
1교시 AI와 윤리, 그리고 해결을 위한 교육  
2교시 프로그래밍 활용 파이게임 만들기와 AI활용 영상만들기  
<https://www.youtube.com/watch?v=7nddZ2zl4G0>



4일차

**5일차**  
1교시 AI로 데이터를 시각화 및 분석해 보자 (엔트리, 오렌지3)  
2교시 AI로 디바이스를 제어해 보자  
<https://www.youtube.com/watch?v=mPmHxEbKQPw>



5일차

**6일차**  
1교시 ChatGPT 활용하기 3 (PPT, 엑셀 자동으로 만들기)  
<https://www.youtube.com/watch?v=cW2umRztNHQ>



6일차

**|총괄·기획|**

정덕근 인천광역시교육청 AI융합교육과장  
정미란 인천광역시교육청 AI융합교육과 장학관  
손보경 인천광역시교육청 AI융합교육과 장학사

**|교육과정 개발 및 내용 검토|**

김석전 송도중학교 교사/인하대학교 겸임교수  
김세호 광성중학교 교사/인하대학교 겸임교수

**|집필진|**

1. AI 이론  
김석전(정보) 송도중학교  
이완범(정보) 인천과학예술영재학교  
강영은(수학) 계산고등학교  
홍순기(역사) 부평고등학교

**2. AI 앱 제작실습**

장희욱(과학) 인천초은중학교  
김세호(정보) 광성중학교  
박현주(수학) 삼산중학교

**3. AI 활용**

정영도(과학) 인하대학교사범대학부속중학교  
김민구(정보) 대건고등학교  
장은혜(과학) 백석고등학교  
서강식(수학) 간석여자중학교

