

학교에서 만나는



인공지능 수업





학교에서 만나는
인공지능 수업

고등학교 기초





들어가는 글

인공지능 기술의 발달로 급변하는 사회를 바라보는 우리는 두려움과 기대가 혼재된 마음으로 미래를 준비하고 있습니다. 그러한 첫 걸음으로 2020년 5월, 교육부와 한국과학창의재단은 2015개정 국가교육과정 고등학교 과정에 ‘인공지능 기초’ 과목을 신설함으로써, 처음으로 인공지능 교육을 공교육에 실현하게 되었습니다.

본 교재는 ‘인공지능 기초’ 과목이 개설되기 전에 인공지능에 대한 일부 핵심 내용을 교과 및 비교과에 선적용해 볼 수 있도록 모듈형으로 구성되어 있습니다. 각 모듈은 ‘인공지능 기초’ 과목에서 가장 필수적이고 핵심적인 개념이라고 할 수 있는 다섯 가지 요소를 선별하였으며, 이론적 내용을 최소화하고 체험과 실습을 중심으로 구성하였습니다.

본 교재의 내용은 다소 어렵게 느껴질 수 있는 인공지능의 개념과 원리에 대해, 학생 누구나 재미있고 흥미롭게 배울 수 있도록 쉬운 내용으로 집필하였습니다. 또한 학생들이 진로와 적성에 상관없이 미래를 살아가는 데 필요한 기본 소양으로서 인공지능을 이해할 수 있도록 다양한 주제의 예시를 활용하였습니다.

따라서 ‘인공지능 기초’ 과목을 개설하지 못한 학교에서도 ‘정보’ 과목에서 부분적으로 연계하여 필요에 따라 확장 수업의 형태로 활용할 수도 있고, 타교과에서도 유사한 주제를 다루는 경우에 모듈별로 적용이 가능할 것으로 기대하고 있습니다.

끝으로 코로나19의 어려움에도 불구하고 수차례 반복적 회의를 통해 본 교재를 제작하기 위해 노력해 주신 모든 분들께 감사드립니다.

2021년 2월

집필진 일동

이 책의 구성과 특징

1단계

핵심 개념 알기

각 영역의 주요 핵심 개념을 탐색하고, 관련된 활동 주제를 미리 살펴본다.

2단계

창의융합적 사고로 해결하기

각 활동 과제를 해결하기 위한 배경 지식을 바탕으로 주어진 문제 상황을 이해하고, 창의융합적으로 문제를 해결하는 과정에서 자기주도적 학습 능력을 기른다.

3단계

공감 나누기

모든 활동을 마친 후 배운 내용을 정리하며 서로의 생각을 공유한다.

1단계

핵심 개념 알기

1단계 핵심 개념 알기

인식

인공지능은 어떻게 세상을 소통할까?

저능 에이전트는 센서로 환경을 인식하여 세상의 상호작용한다.

1. 센서와 인식

Q1.1 센서는 어떻게 주변 환경을 인식할까?

센서는 저능 에이전트 존재하는 물리량 또는 대상물의 정보를 측정해 전기 신호로 바꾸는 장치이다. 센서에 따라 기능을 추가하여 인식하고 상황을 판단하는 것을 감지라고 한다.

스마트폰 센서의 종류에는 동작, 위치, 환경 센서 등이 있다.

유형	기능
동작	움직임 센서는 사용자 움직임과 걸음걸이, 가속도와 중력, 회전 속도, 회전 방향 및 방향 정보를 측정한다.
위치	위치 센서는 기기의 물리적인 위치를 확인하는 데 사용된다. 자기가 센서와 가속도계를 활용해서 내장된 기능을 제공하고, 기기 위치를 확인한다.
환경	환경 센서는 습도, 온도, 압력, 운동 등을 알려 주는 4개의 센서이다. 온도 센서는 기기 내부의 온도를 측정하며, 마이크로폰은 외부의 음향 정보를 수집한다.

2. 컴퓨터 비전

Q1.2 인공지능은 어떻게 세상을 볼까?

컴퓨터 비전(Computer Vision)은 인간의 시각처럼 컴퓨터 기기가 자율적으로 할 수 있도록 만든 분야이며, 인간과 다른 시각 기능을 부여하거나 각기 다른 능력으로 새로운 시야를 만들고, 인간의 한계를 보완하는 기술이다. 그래서 자율주행뿐만 아니라 컴퓨터의 눈이 필요한 여러 분야에서 활용된다.



40 | 1. 인식

"저능 에이전트는 센서, 시각, 청각 등의 인식을 통하여 세상을 상호작용한다."

3. 음성 인식

Q1.3 인공지능은 어떻게 소리를 인식할까?

음성 인식(Speech Recognition)이란 사람이 말하는 언어를 컴퓨터가 해석해 그 내용을 문자로 변환하는 것을 말한다.



▲ 인공지능 ASMR 영상 인식 과정
음성 인식 기술은 전처리와 인식의 두 가지 단계로 구성된다. 전처리는 사용자가 발생한 음향으로부터 인식할 대상이 되는 구간 찾기, 잡음 제거, 소리의 특징을 추출하는 과정이다. 인식은 음성에서 특징을 추출하여 패턴을 분석하고 언어를 식별하는 과정을 포함한다.

4. 언어 이해

Q1.4 인공지능은 어떻게 언어를 이해하고 처리할까?

인공지능이 사람의 말을 이해하고, 처리할 수 있는 능력의 기본은 자연어 처리(Natural Language Processing) 기술이다. 자연어 처리는 기계된 컴퓨터가 인간의 언어를 이해할 수 있도록 하는 것으로, 자연어 이해와 자연어 생성의 두 가지로 구성된다. 자연어 이해는 의미의 역할과 같은 문맥에서 의미 해석을 위해 기계가 인간의 언어를 이해하고 분석하는 데 도움을 준다. 자연어 생성은 컴퓨터가 데이터를 자연어 형태로 변환하는 번역기 역할을 한다. 이러한 자연어 처리 기술이 적용된 대표적인 사례로는 기계 번역, 챗봇, 스피킹 장치, 감정 분석, 워드 클라우드, 맞춤형 추천 등이 있다.

활동 1.1 보기

핵심 개념 관련된 활동은 어떤 내용으로 진행할까요?

활동 1

· 센서는 어떻게 주변 환경을 인식할까?
· 다양한 스마트폰 센서 활용하기
· 다양한 센서 활용 아이디어 발표하기

활동 2

· 인공지능은 어떻게 세상을 볼까?
· 타자용 인식에서 사물 인식 체계화하기
· 이미지 인식 알고리즘을 만들기
· TensorFlow로 손글씨 인식하기
· 컴퓨터 비전의 한계

활동 3

· 인공지능은 어떻게 소리를 인식할까?
· 타자용 인식에서 사람 목소리 구분하기
· 음성 인식으로 글자 인식하기
· 인공지능 스피커와 대화하기

활동 4

· 인공지능은 어떻게 언어를 이해할까?
· 워드 클라우드로 말의 주제를 파악하기
· 번역기 체험하기
· 워드 클라우드로 말의 주제를 파악하기

단계별 학습하기 | 41

2단계

창의융합적 사고로 해결하기

#2 인공지능은 어떻게 이미지를 인식할까?

→ 이미지 인식과 컴퓨터 비전의 원리를 이해하고, 이미지 인식의 한계에 대해 설명할 수 있다.

1 문제 이해하기

사례 두 가지를 읽고, 인공지능이 이미지 인식의 원리와 한계를 생각해 보자.

사례 1



과일 농가에서 과일을 등급별로 분류하는 일에도 인공지능이 적용되고 있다. 딥러닝과 컴퓨터 비전을 활용한 '농산물 자동 선별기'는 농산물의 종류와 무게를 인식해 자동으로 과일의 무게를 분류해 준다. 농산물의 모양과 색깔, 크기에 따라 다양한 등급으로 선별하는 것을 활용, 농산물에 생긴 흙집과 병 등을 감지해 상용성이 떨어지는 것을 자동으로 골라낸다. 이처럼 기계에 상용성을 인식하고 자동으로 선별하는 과정에서 재물 손상을 줄일 수 있다.

사례 2

고속도로에서 자율주행 보조로 지원하던 자율주행자동차가 판매된 차선에서 좌회전 하는 차량 보행자가 보행자를 인식하지 못해 교통사고가 일어났다. 자율주행자동차는 화면에 대형 보행자 보행자의 옆면 아랫부분과 정면 충돌하였는데, 그 이유는 당시 날씨가 매우 화창해 자율주행 제어 장치나 센서가 모두 위층의 보행자의 모습을 자동으로 인식하지 못해 사고가 난 것으로 보고 있다.



과일 분류기로 어떻게 과일을 분류할 수 있을까? 그리고 이러한 기능의 한계는 무엇일지 생각해 보자.

50 p. 1면

3단계

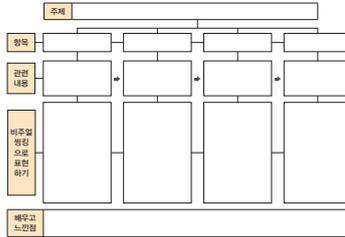
공감 나누기

3단계 공감 나누기

실용하기

인식 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

1 배운 내용을 비주얼 씬킹으로 표현하고, 친구들과 비교해 보자.



[작성 방법] 1. 항목에는 인식의 4가지 활동을 작성한다.
2. 관련 항목에는 주요 학습 요소를 작성한다.
3. 알게 된 사실을 글과 그림으로 표현한다.



80 p. 1면

무엇을 알아야 할까?

배경 지식

컴퓨터 비전과 이미지 인식

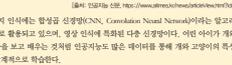
컴퓨터 비전은 이미지나 영상과 같은 시각 데이터를 컴퓨터가 이해할 수 있게끔 처리하는 분야이다. 인간의 시각이 할 수 있는 일을 보조하거나 대체하는 자율적인 시스템을 만드는 것을 목표로 한다. 컴퓨터 비전은 글자와 같은 2차원 화제 패턴 인식의 2차원 도형이나 사물을 분류하는 것, 3차원 공간에서 물체를 인식하고 추적하는 것 등 다양하다. 컴퓨터 비전은 무인택시, 자율주행자동차, 로봇, 생활 속 카메라와 스마트 기기 등에 활용된다.

이미지 인식은 이미지의 정보를 식별하는 기술로, 컴퓨터 비전의 핵심 기술 중 하나이다. 이미지에서 포함된 객체를 인식하고 분류하는 것이 대표적이다. 컴퓨터 비전에서 이미지 캡처링(Image Capturing)이란 컴퓨터가 사물을 보고 적절한 설명을 자동으로 붙이는 기술이다. 캡처링 알고리즘은 이미지를 입력하면 이미지를 설명하는 문장을 출력한다.



▲ 자율주행자동차의 안전 운전 ▲ 인공지능의 역할

인공지능은 사물 분류 이전에 특정 보드를 읽고 있는 날씨, 오른쪽 그림에는 '달란 속 보리'에 대한 사진이라는 정보를 준다. 인공지능 기술의 활용도 다양해 이미지를 인식 및 분석해 관련성을 자연어로 해석해 주는 것이 가능해진다.



이미지 인식에는 딥러닝을 신경망(ANN, Convolutional Neural Network)이라는 알고리즘이 대표적으로 활용되고 있으며, 영상 인식에 특화된 다층 신경망이다. 어떤 사람이 개와 고양이 의 모습을 보고 배우는 것처럼 인공지능도 많은 데이터를 통해 고양이 특징을 추출하여 단계적으로 학습한다.

진짜 보충학습 자료 51

2 프로젝트를 진행하면서 인간과 인공지능이 인식의 관점에서 어떻게 다른지 비교하여 작성해 보자.

인간	인공지능

3 앞으로 인간과 인공지능이 공존하는 미래의 모습을 친구와 함께 이야기 쓰기로 작성해 보자.

주제: 인공지능과 함께 하는 미래의 생활에서 우리 인간은 어떤 노력을 해야 할까?

(오동천1)
(오동천2)
(오동천3)
(오동천4)
(오동천1)
(오동천2)



진짜 보충학습 자료 81

I 지능 에이전트

1단계	핵심 개념 알기: 지능 에이전트	10
2단계	창의융합적 사고로 해결하기	12
활동 1	인공지능이란 무엇일까?	12
활동 2	나의 직업은 어떻게 변할까?	18
활동 3	나에게 필요한 지능 에이전트는?	24
3단계	공감 나누기	36

II 인식

1단계	핵심 개념 알기: 인식	40
2단계	창의융합적 사고로 해결하기	42
활동 1	센서는 어떻게 주변 환경을 인식할까?	42
활동 2	인공지능은 어떻게 이미지를 인식할까?	50
활동 3	인공지능은 어떻게 음성을 인식할까?	58
활동 4	인공지능은 어떻게 사람의 말을 이해할까?	68
3단계	공감 나누기	80

III 데이터

1단계	핵심 개념 알기: 데이터 이해와 활용	84
2단계	창의융합적 사고로 해결하기	86
활동 1	사람의 기대 수명은 얼마나 될까?	86
활동 2	다른 사람들은 인터넷으로 무엇을 할까?	94
활동 3	올 겨울 유행할 외투는 무엇일까?	104
활동 4	유명 롤러코스터를 살펴볼까?	114
활동 5	남녀공학 여부에 따른 나의 성적 평가는?	124
3단계	공감 나누기	138

IV 기계학습

1단계	핵심 개념 알기: 기계학습과 인공지능 프로젝트 사이클	142
2단계	창의융합적 사고로 해결하기	144
활동 1	기계학습으로 무엇을 할 수 있을까?	144
활동 2	인공지능으로 해결할 수 있는 문제를 찾아볼까?	152
활동 3	기계학습 모델이 학습할 데이터는 어떻게 모을까?	158
활동 4	나의 모델은 얼마나 똑똑해졌을까?	168
3단계	공감 나누기	184

V 인공지능 윤리

1단계	핵심 개념 알기: 인공지능의 윤리적 이슈	188
2단계	창의융합적 사고로 해결하기	190
활동 1	편향성은 어떻게 생기는 걸까?	190
활동 2	인공지능의 선택은? 당신의 선택은?	200
3단계	공감 나누기	210

I 지능 에이전트

1 단계 핵심 개념 알기

- 인공지능과 직업
- 지능 에이전트
- 인공지능과 지능 에이전트의 관계



2 단계 창의융합적 사고로 해결하기

- 활동 1 인공지능이란 무엇일까?
- 활동 2 나의 직업은 어떻게 변할까?
- 활동 3 나에게 필요한 지능 에이전트는?



3 단계 공감 나누기

- 프로젝트 정리하기

④ 지능 에이전트 단원에서는 무엇을 배울까요?

지능 에이전트의 개념을 이해하고 인공지능과 지능 에이전트의 관계를 탐색하며, 인공지능이 직업에 미칠 영향과 내 직업의 미래를 예측해 봅니다. 그리고 이를 바탕으로 나만의 지능 에이전트를 설계해 봅니다.





1단계 핵심 개념 알기

지능 에이전트

✔ 인공지능과 지능 에이전트는 어떤 관계일까?

우리가 생활 속에서 경험하는 인공지능이 실제 현실 세계에 작용하려면 에이전트의 모습으로 구현된다. 인공지능과 지능 에이전트의 관계를 이해하면 우리 생활에 인공지능이 어떻게 활용되고 있는지 알 수 있다.

1 인공지능

Q 개념 탐색 인공지능이란 무엇일까?

인공지능은 인간이 만든 지능이라는 의미로 사용되지만 튜링이 제안한 ‘튜링 테스트’에서 시작하여 현재는 ‘인간의 지능을 모방하여 인간과 비슷하게 생각하고 행동하는 기계를 만드는 것’이라고 볼 수 있다. 인간과 비슷한 행동을 한다는 기준이 모호해서 컴퓨터 과학에서는 ‘합리적인 생각과 합리적인 행위를 할 수 있는 기계를 만드는 것’이라고 설명한다.

1956년 다투머스 회의에서 존 매카시는 인공지능을 ‘지능형 기계를 만드는 과학 및 공학’으로 정의하였고, 이후 1959년에 아서 사무엘이 인공지능의 핵심 분야인 기계 학습(machine learning)을 ‘명시적인 프로그래밍 없이도 컴퓨터가 스스로 학습할 수 있도록 하는 연구 분야’로 정의하였다.

따라서 인공지능은 ‘인간의 지능적인 행위, 즉 의사결정, 판단, 추론, 예측 등을 컴퓨터가 수행하도록 만드는 기술 또는 이를 연구하는 학문 분야’라고 설명할 수 있다.

2 지능 에이전트

Q 개념 탐색 지능 에이전트란 무엇일까?

에이전트는 ‘대리인’, ‘대행자’란 뜻으로 컴퓨터 분야에서는 사람 대신 업무를 처리하는 소프트웨어 또는 하드웨어를 말한다. 지능 에이전트란 ‘사람의 지능적인 일을 대신할 수 있는 소프트웨어나 하드웨어’를 의미하며, 주변 환경에 스스로 적응할 수 있는 형태로 구현된다.

에이전트 관점에서 인간과 로봇, 소프트웨어를 비교하면 다음과 같다.

- 인간 에이전트: 오감으로 정보를 받아들이고, 팔다리 등을 움직여 일을 수행한다.
- 로봇 에이전트: 카메라, 거리 센서 등을 통해 정보를 받아들이고, 다양한 모터를 움직여 일을 수행한다.
- 소프트웨어 에이전트: 키 입력, 파일 내용, 네트워크 신호 등을 통해 정보를 받아들이고, 화면 표시, 파일 기록, 네트워크 신호 등을 통해 일을 수행한다.

지능 에이전트의 속성에는 자율성, 사회성, 적응성 등이 있어요.



“에이전트는 감지기와 작동기를 통해서 환경과 상호작용한다.”

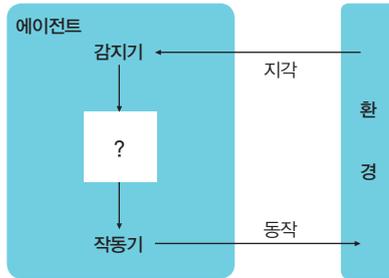
3 인공지능과 지능 에이전트

Q 개념 탐색 인공지능과 지능 에이전트의 관계를 알아보자.

인공지능은 인간처럼 합리적으로 판단하고 행동하는 것을 컴퓨터로 구현하는 ‘기술 또는 시스템’이므로 실생활에서 활용하기 위해서는 지능 에이전트로 구현되어야 한다.

1 지능 에이전트

지능 에이전트가 합리적으로 작동하려면 어떤 환경에서, 어떤 감지기를 이용해서 지각하고, 어떤 작동기를 이용해서 동작을 수행할지, 그에 따른 성과는 어떻게 측정할지를 고려해야 한다.



▲ 지능 에이전트의 요소와 작동 방식

[출처: 스투어트 러셀, 피터 노빅, (2016). 인공지능: 현대적 접근방식(제3판)]

- 인공지능은 이미지를 자동으로 인식하거나 목소리를 자동으로 판별해주는 등의 소프트웨어나 하드웨어 형태로 나타난다.
- 지능 에이전트는 주변 환경을 스스로 인식하고 적응하며 행위할 수 있도록 구현된다. 즉, 감지기를 통해서 환경을 지각하고, 작동기를 통해서 환경에 따라 동작하는 과정을 수행하는 것으로 지각, 판단, 행동의 과정을 반복하면서 합리적으로 행동한다.
- 인공지능은 지능 에이전트를 더 합리적이고 지능적으로 만들어 주는 기술 또는 시스템이라고 볼 수 있다.

2 인공지능 적용 사례

인공지능이 인간과 비슷한 수준의 합리적인 행동을 할 수 있다면 사람을 대신하여 판단하고, 행동할 수 있다. 인공지능의 역할은 어디까지 가능할까? 인공지능은 문제 해결, 추론, 학습, 인식과 같은 분야에서 좋은 성능을 보이고 있다.

- 문제 해결: 체스, 바둑, 스도쿠, 내비게이션, 로봇팔 등
- 추론: 전문가 시스템, 챗봇, 로봇 등
- 학습: 기계학습, 예측 시스템, 데이터 마이닝 등
- 인식: 컴퓨터 비전, 번역, 음성 인식 등

활동 미리 보기

핵심 개념과 관련된 활동은 어떤 내용으로 전개될까요?

활동 1

인공지능이란 무엇일까?

- 튜링 테스트 활동하기
- 인공지능 특성 탐색하기

활동 2

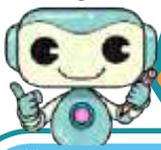
나의 직업은 어떻게 변할까?

- 인공지능 서비스 체험하기
- 인공지능과 미래 직업 변화 조사하기
- 나의 희망 직업 변화 예측해 보기

활동 3

나에게 필요한 지능 에이전트는?

- 지능 에이전트 체험하기
- 지능 에이전트와 협업하며 작동 원리 분석하기
- 지능 에이전트 사례와 역할 보고서 작성하기
- 나에게 필요한 지능 에이전트 구상하기



2단계 창의융합적 사고로 해결하기

활동 1 인공지능이란 무엇일까?

- ➔ 인공지능의 개념을 이해하고 인공지능인 것과 인공지능이 아닌 것을 구분하여 설명할 수 있다.

1 문제 이해하기

A와 B를 읽고, 사람이 지은 시와 인공지능이 지은 시를 구별해 보자.

A

육첩방은 백골이 따라와
 좋아하지 방에 누웠다.
 한 밖에 밤비가 속살거리는데
 시인이란 너희의 바람이

B

우물 속에는 달이 밝고 구름이 흐르고
 하늘이 펼치고 파아란 바람이 불고
 가을이 있고 추억처럼 사나이가 있습니다.

인공지능이 발전하면서 인간의 고유한 창작 분야라고만 생각했던 문학, 음악, 미술과 같은 예술 영역까지 인공지능이 도전하는 사례가 늘고 있다. 따라서 최근에는 인공지능 작가, 인공지능 음악가, 인공지능 화가와 같은 지능 에이전트가 작품을 출품하고 전시회를 여는 사례 역시 늘어나고 있다.

위에서 살펴본 두 개의 시 중 B시는 윤동주 시인의 작품인 '자화상'의 일부이며, A시는 2019년 문화체육관광부와 한국콘텐츠진흥원이 진행한 'Impact X' 무대에서 소개된 작품이다. 이 작품은 윤동주 시인의 작품을 학습한 인공지능이 윤동주 시인의 시풍으로 새로운 시를 창작하는 'AIX 윤동주' 프로젝트의 결과물이다.

인공지능과 인공지능이 아닌 것의 특징을 생각해 보자.

- 인공지능은 _____ 이다.
- 인공지능이 아닌 것은 _____ 이다.



무엇을 알아야 할까?

배경 지식

인공지능의 특성

인공지능은 컴퓨터가 인간과 같이 지능적인 행위를 할 수 있도록 만드는 기술을 의미한다.

인공지능인지 아닌지를 구별하려면 우선 인공지능의 특성을 파악해야 한다. 인공지능은 인간이 가지고 있는 지능을 흉내내어 만들었기 때문에 인간 지능 혹은 자연 지능의 특성을 반영하는 지능 에이전트 형태로 구현되었다.

인공지능을 탑재한 지능 에이전트는 다음과 같은 행동 특성을 가진다.

- 자율성: 사용자나 다른 프로그램의 직접적인 지시나 간섭 없이 스스로 판단하여 작업을 수행하는 능력
- 사회성: 지능 에이전트가 통신 언어를 사용하여 사람과 다른 에이전트와 상호작용하는 능력
- 적응성: 환경의 변화에 유연하게 대응하고, 경험을 통해 학습하며 목표를 달성하도록 적응하는 능력



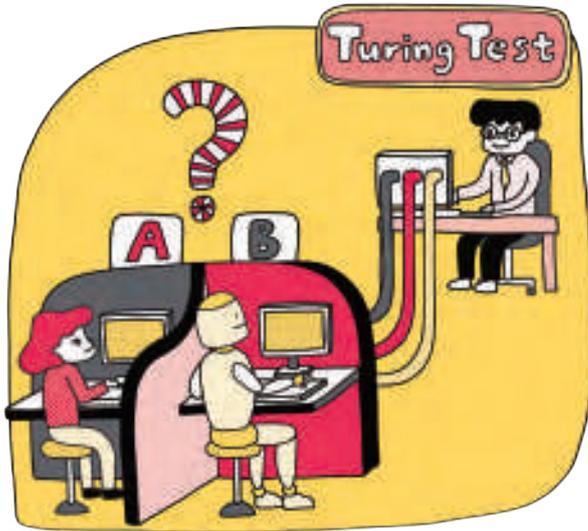
이 외에도 에이전트의 구성 원리에 따라 센서로부터 입력을 받고 행동을 수행하여 환경을 변화시키는 상황성을 가지며, 합리성, 협동성, 능동성, 목표 지향성 등의 특성을 가진다.

이러한 특성에 따라 우리가 많이 사용하는 SNS 플랫폼이나 온라인 공유 플랫폼에도 인공지능이 지능 에이전트 형태로 작동하고 있는 것을 확인할 수 있다. 동영상 공유 사이트에서 사용자의 검색어 내역을 학습하여 이를 기반으로 동영상 콘텐츠를 추천하거나, 인터넷 쇼핑몰 사용자가 물건을 구매한 내역에 따라 다양한 쇼핑 아이템을 추천하는 것을 예로 들 수 있다.

2 문제 상황 체험하기

튜링 테스트를 통해 인공지능을 알아보자.

- 1** 튜링 테스트는 인공지능의 개념을 처음 제안한 앨런 튜링이 만든 테스트이다.
아래 그림을 보고 튜링 테스트가 무엇인지 살펴보자.



A방과 B방에 각각 인간과 기계(인공지능)를 배치한다. 판정자는 이들의 정체를 모르는 채로 양쪽과 텍스트로 대화를 나눈다. 대화를 나눈 뒤, 판정자가 기계와 인간을 정확히 판정하지 못하면 기계가 지능을 가졌다고 판단한다.

2 튜링 테스트 활동을 해 보자.

- ① 학급에서 4명의 지원자를 선발한 뒤, 역할을 부여한다(기계 2명, 인간 2명).
- ② 인간과 기계 역할을 맡은 학생은 비밀리에 짝을 지어 교실 앞쪽과 뒤쪽에 앉는다.
- ③ 나머지 학생들은 다음과 같이 질문 목록과 개인 질문 1~2개를 만들어서 교실 앞이나 뒤로 가서 질문한다.

질문 목록	사전 답변	개인 질문
<ul style="list-style-type: none"> • 당신은 컴퓨터인가요? • 3, 6, 9, 12, 15 다음에 올 숫자는? • 34×27의 값은? • 4의 제곱근은? • $3456 + 5677$의 값은? • 오늘은 무슨 요일인가요? • 취미가 무엇인가요? • 독서를 좋아하나요? • 좋아하는 게임은 무엇인가요? 	<ul style="list-style-type: none"> • 아니요. • 18 • 918 • 2 • 9133 • 해당 요일로 답변 • 취미는 웹서핑이에요. • 네. • 마인크래프트입니다. • 이외의 질문에는 다음과 같이 답변하기: 저는 잘 모르겠어요. 당신은 어떻게 생각하나요? 	<ul style="list-style-type: none"> • • •

- ④ 각 역할을 맡은 학생들은 사전에 주어진 답변 방법대로만 답변한다.
- ⑤ 질문을 주고받은 뒤 어느 쪽이 인간인지 생각해 보고, 그 이유를 적어보자.

인간인 쪽	이유

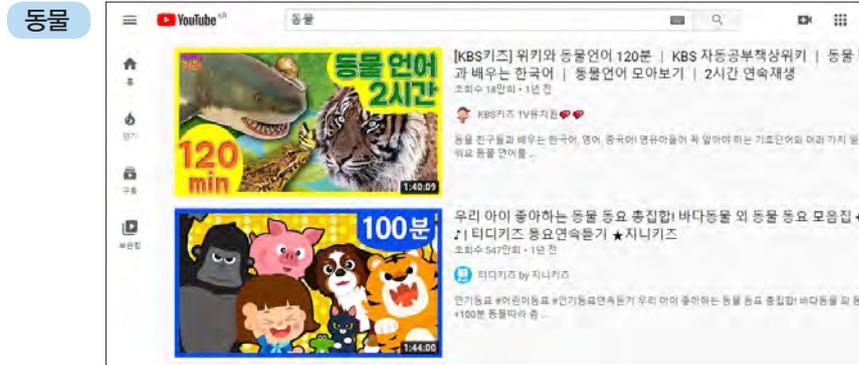


3 인공지능의 특성 찾아보기

인공지능이 적용된 서비스를 통해 인공지능의 특성을 살펴보자.

1 유튜브에 적용된 인공지능을 찾아보자.

- ① 유튜브에 접속한다.
- ② '동물', '식물'을 검색해 본다.



- ③ 각자의 검색 결과를 비교해 본다.

2 유튜브에 적용된 인공지능을 다음 네 가지 관점에서 생각해 보자.

의사결정	예측	인식	학습
------	----	----	----

3 유튜브에 인공지능이 적용되어 검색 결과가 다르게 나타난 것처럼 인공지능이 적용된 서비스를 찾아보고, 어떤 특성이 있는지 생각해 보자.

4 정리하기

1 인공지능의 개념을 정리해 보자.

• 인공지능이란 _____ 이다.

• 인공지능의 특성에는 _____ , _____ , _____ 등이 있다.

_____ 이고,

_____ 이며,

_____ 이다.

2 즐거운 인공지능 생활을 위한 문장 하나를 써 보자.

• 인공지능은 (_____) 이다.

3 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
인공지능 개념	인공지능의 개념을 설명할 수 있는가?					
인공지능 특성	인공지능의 특성을 설명할 수 있는가?					
협력	튜링 테스트 활동에 적극 참여하였는가?					

인공지능의 개념을 명확하게 정의하기는 힘들다. 컴퓨터의 발달과 인공지능 연구 시기에 따라 초점과 범위가 변하고 있기 때문이다. 인공지능을 연구하는 학자들 사이에서도 접근 방법에 따라 인공지능을 다르게 정의하고 있다.

인공지능이라는 단어로 그 의미를 유추해 보면 ‘인공’은 사람이 만들었다는 의미이고, ‘지능’은 지적 활동의 능력이라는 의미이다. 사람이 만들었다는 것은 명확히 구분하고 정의할 수 있으나 ‘지적 활동의 능력’은 명확하게 정의하기 어렵고 과학자들도 다른 견해를 보인다. 인간의 지적 활동은 주로 지식 습득, 논리적 추론, 추상적 사고, 계획과 실행, 독창성, 문제 해결, 정서 지식 등과 관련되어 있다.

앞서 살펴본 튜링 테스트에서도 어느 정도의 자연스러움을 보이면 지능이 있다고 판별할지 그 비율을 정하는 것에서도 전문가들마다 의견이 다르다. 또한 지능적이라고 할 때, 지적 능력뿐만 아니라 일을 처리하는 방식도 중요하게 다루어야 한다는 주장도 있다. 예를 들어 바둑을 배울 때 데이터만 가지고 학습하는 것이 아니라 인간이 바둑을 배우는 방식으로 배울 수 있는지도 중요하다는 것이다. 지능을 연구하는 분야는 철학, 언어학, 심리학, 교육학, 신경과학, 뇌 과학 등이 있는데 아직 ‘지능’의 정의와 메커니즘을 명확하게 밝혀내지 못했다.

인공지능은 ‘컴퓨터가 생각할 수 있을까?’ ‘컴퓨터가 스스로 학습할 수 있을까?’와 같은 질문에 답하기 위해 발전했기 때문에 인간 지능만큼의 결과를 내면 지능이 있다고 받아들이자는 입장과 결과만 가지고 지능이 있다고 여긴다면 인간이 가진 지능의 가치를 너무 낮게 평가한 것이라는 주장이 있다. 전자가 튜링의 견해라면 후자는 철학자 존 설(John Searle)의 의견으로, 지능적인 결과를 보이더라도 지능적인 행위를 진정으로 이해하지 못하기 때문에 지능이 있다고 볼 수 없다는 주장을 펼쳤다.

현대 컴퓨터 과학 연구에서는 이런 논쟁을 줄이기 위해 인공지능의 정의를 다음과 같은 차원으로 설명하고 있다.

← 강인공지능

약인공지능 →

구분	인간다움(Human-like)	합리적(rational)
사고	<ul style="list-style-type: none"> 인간처럼 사고한다. 의사결정, 문제 해결과 같은 인간의 인지 능력을 자동화시킨다. 	<ul style="list-style-type: none"> 합리적으로 사고한다. 인지와 추론, 행위를 가능하게 하는 계산 모델을 만들어서 작동하게 한다.
행동	<ul style="list-style-type: none"> 인간처럼 행동한다. 사람이 지능적으로 수행하는 것을 컴퓨터도 수행할 수 있도록 자동화시킨다. 	<ul style="list-style-type: none"> 합리적으로 행동한다. 지능적인 행위를 할 수 있는 지능 에이전트(Intelligent agents)를 설계하고 구현한다.

[출처: 스텐튼 러셀, 피터 노빅.(2016). 인공지능: 현대적 접근방식(제3판)]

강인공지능은 다양한 분야에 보편적으로 활용할 수 있도록 사람처럼 사고하고 행동하는 단계이고, 약인공지능은 지능이 있는 것처럼 프로그래밍된 것으로 특정 분야에서 활용되는 수준의 인공지능을 말한다. 지금까지 개발된 인공지능은 약인공지능에 해당되며, 앞으로 과학 기술의 발전에 따라 인공지능의 개념이 어떻게 변할지 살펴볼 필요가 있다.

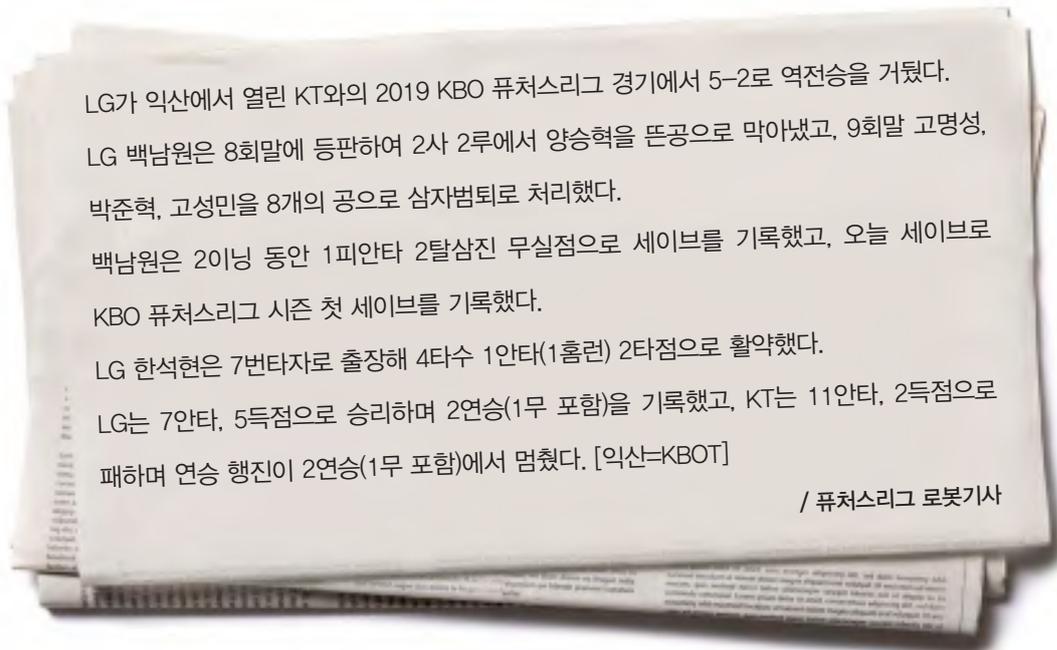
활동 2

나의 직업은 어떻게 변할까?

➔ 인공지능의 발전에 따라 직업이 어떻게 변화할지 예측할 수 있다.

1 문제 이해하기

로봇 기자 케이봇(KBOT)이 작성한 기사를 읽고, 사람이 쓴 기사와 어떤 차이점이 있는지 살펴보자.



사람 대신 로봇 기자가 신문 기사를 쓴다면 우리는 어떤 역할을 할 수 있을지 생각해 보자.

무엇을 알아야 할까?

배경 지식

인공지능의 활용 분야

인공지능은 사회와 산업의 다양한 분야에서 활용되고 있다. 인공지능 국가 전략에 나타난 다양한 서비스를 통해 각 분야별로 인공지능을 어떻게 활용하고 있는지 확인할 수 있다.



[복지] 어르신의 말동무이자 보호자, 김포시 챗봇 '다솨이'

- AI 케어 로봇 '다솨이'는 어르신이 30분 이상 말이 없으면 먼저 말을 걸고 5시간 이상 움직임이 없을시 보호자·생활 관리사를 자동으로 연결한다.



[교육] 학생과 영어로 대화하는 서울 교육청 인공지능 영어 보조교사

- 인공지능 영어교사가 학생과 영어로 대화를 나누거나 퀴즈를 출제하는 등 영어 교사의 수업을 보조한다.



[농업] 작물에 따라 환경을 제어하는 '플랜티 큐브'

- 지능화 기술을 활용한 컨테이너 농장으로 농장의 규모와 작물 수요에 따라 환경을 제어하여 고품질 작물을 연 최대 13회까지 수확할 수 있다.

[출처: 대한민국 정부, 2019.12. 인공지능 국가전략, 8p]

다음 사례를 통해 인공지능이 다양한 사회 분야에서 활용되고 있다는 점을 확인할 수 있으며, 인공지능의 주요 기술과 적용 가능한 산업 분야를 조사하면 미래 직업 사회 변화를 예측할 수 있다.

기술 \ 분야	건강	모빌리티	금융	온라인 서비스	가정·개인
음성 인식 & 자연어 처리	원격 진료	음성 인식 명령	챗봇	챗봇	개인 비서
이미지 인식	원격 진료	교통 신호 인식	보안	보안	가정 로봇
패턴 인식	개별 건강 패턴	이동패턴	주식 추천	물품 추천	수면 패턴

[출처: Consumers International, Artificial Intelligence: consumer experiences in new technology - RESEARCH AND WORKSHOP REPORT APRIL 2019]

2 문제 상황 체험하기

인공지능이 적용된 서비스를 체험해 보고, 직업에 미칠 영향을 생각해 보자.

1 검색 엔진에서 '파파고'를 검색한 후, 사이트에 접속한다.



2 자신을 소개하는 글을 쓰고, 파파고에 입력하여 어떻게 번역하는지 살펴보자.

안녕하세요?
저는 _____입니다.
_____ 고등학교에 다니는 _____학년 학생입니다.
_____를 좋아하고, _____를 잘합니다.
친구들과 _____를(을) 먹는 것을 좋아합니다.

3 번역 결과를 확인하고 정확도를 평가하여 별을 색칠한 뒤, 친구들과 비교해 보자.

※ 평가도에 따라 별표에 색칠해 보세요.

내가 판단한 정확도	☆☆☆☆☆
_____가 판단한 정확도	☆☆☆☆☆
_____가 판단한 정확도	☆☆☆☆☆
_____가 판단한 정확도	☆☆☆☆☆

3 생각해 보기

인공지능 기술의 발전으로 인해 직업은 어떻게 변화할지 조사해 보자.

1 아래 직업 중에 사라질 확률이 낮은 직업과 높은 직업을 선택해서 적어 보자.

콘크리트공	화가 및 조각가	경리사무원	가수
정육원 및 도축원	사진작가	환경미화원	메이크업 아티스트
고무 및 플라스틱 제품 조립원	작가	택배원	공예원
청원경찰	지휘자·작곡가 및 연주자	과수 작물 재배원	예능 강사
조세 행정 사무원	애니메이터 및 만화가	행정 및 경영 지원 관련 서비스 관리자	패션 디자이너
물품 이동 장비조작원	무용가 및 안무가	주유원	국악 및 전통 예능인
부동산 중개인	감독		

[출처: 한국고용정보원]

사라질 확률이 낮은 직업	사라질 확률이 높은 직업
• 직업: • 이유:	• 직업: • 이유:

2 인공지능 기술의 발전으로 영향을 받게 될 직업군을 조사해 보자.

인지적	기술 대체 가능성 중간	기술 대체 가능성 낮음
육체적	기술 대체 가능성 높음	기술 대체 가능성 낮음
	정해진 업무가 명확한 경우	정해진 업무가 불명확한 경우

3 인공지능의 영향으로 인공지능과 협업 또는 새로 생겨날 직업과 사용 부분을 조사해 보자.

직업	인공지능 사용 부분

4 나의 희망 직업 변화 예측해 보기

1 나의 희망 직업을 적어 보자.

2 나의 희망 직업에서 요구하는 역할과 그 작업 내용을 써 보자.

역할	작업 내용

3 2에서 작성한 역할과 작업 내용에서 인공지능에 의해 처리될 수 있는 정도를 예측해 보고, 어떻게 활용할지 생각해 보자.

작업 내용	인공지능 대체 가능성		
	<input type="checkbox"/> 낮음	<input type="checkbox"/> 보통	<input type="checkbox"/> 높음
	<input type="checkbox"/> 낮음	<input type="checkbox"/> 보통	<input type="checkbox"/> 높음
	<input type="checkbox"/> 낮음	<input type="checkbox"/> 보통	<input type="checkbox"/> 높음
	<input type="checkbox"/> 낮음	<input type="checkbox"/> 보통	<input type="checkbox"/> 높음

자신의 예측 결과를 친구들과 공유하고 비교해 보세요.



4 나의 희망 직업이 수십 년 후 어떻게 변화할지 예측해 보자.

희망 직업명	5년 후	10년 후	20년 후	30년 후
	유지된다 / 변한다	유지된다 / 변한다	유지된다 / 변한다	유지된다 / 변한다
	변하는 부분 ()	변하는 부분 ()	변하는 부분 ()	변하는 부분 ()

5 나의 희망 직업 예측 시나리오를 작성해 보자.

글, 4컷 만화, 그림 등 다양한 방법으로 표현해 보세요.



5 정리 및 평가하기

1 인공지능으로 인한 나의 희망 직업 변화 예측 활동을 통해 알게 된 점을 써 보자.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
직업의 변화	인공지능의 발전에 따라 직업이 어떻게 변화할지 설명할 수 있는가?					
나의 직업 예측	인공지능을 나의 희망 직업에 어떻게 활용할 수 있는지 설명할 수 있는가?					
인공지능의 활용	인공지능을 자신의 업무에 적절하게 활용하려는 태도가 있는가?					

활동 3

나에게 필요한 지능 에이전트는?

- ▶ 지능 에이전트의 종류와 환경을 분석할 수 있다.
- ▶ 나에게 필요한 지능 에이전트를 구상할 수 있다.

1 문제 이해하기

인공지능 개인 비서는 내가 알고 싶은 정보를 얼마나 정확하게 찾아 알려 줄 수 있을까? 그리고 나와 얼마나 자연스럽게 대화를 나눌 수 있을까?

스마트폰에서 구글 어시스턴트나 시리, 빅스비 등의 인공지능 개인 비서를 실행해서 다음과 같은 질문을 하면 어떻게 대답하는지 확인해 보자.



인공지능 개인 비서와의 대화가 만족스러웠는지 평가해 보자.

무엇을 알아야 할까?

배경 지식

지능 에이전트의 종류와 환경 분석

지능 에이전트가 어떻게 동작하는지 이해하기 위해서는 성과를 측정하고, 환경, 감지기, 작동기와 같은 요소들을 분석하는 것이 좋다. 자율주행자동차를 예로 들면 다음과 같은 요소들로 분석할 수 있다.



- 성과 측정**
 안전, 빠름, 법규 준수, 편안한 이동 등
- 환경**
 도로, 다른 차량, 운전자 등
- 감지기**
 카메라, 속도계, GPS, 주행거리계, 스크린터치 등
- 작동기**
 운전대, 가속기, 제동기, 신호, 경적, 제어 화면 등

지능 에이전트가 합리적으로 동작하려면 지능 에이전트가 일하는 환경 안에서 에이전트의 행동을 평가하는 성과 측정 요소가 필요하다. 우선 과제를 수행하는 환경이 어떤지 파악해야 하는데, 이때 외부 환경으로부터 데이터와 정보를 받아들이는 감지기, 목적지까지 안전하게 이동하기 위한 작동기가 필수적으로 필요하다.

이러한 관점에서 구분한 여러 지능 에이전트의 종류는 다음과 같다.

에이전트 종류	성과 측정	환경	감지기	작동기
물품 정리 로봇	물품이 정확히 정리된 비율	물류 창고, 정리 박스 등	카메라, 관절 각도 감지기, 거리 센서 등	관절로 연결된 팔과 손, 바퀴 등
인공지능 영어 시험 프로그램	학생의 점수	사용 학생들, 시험 기관 등	정답을 입력하는 키보드와 마우스, 마이크 등	연습 문제, 추천 코스, 정오답 등을 표시하는 화면
암 진단 시스템	건강한 환자, 비용 절감	환자, 병원, 의료진	증상, 소견, 환자의 응답을 입력하는 키보드와 마우스	질문, 검사, 진단, 치료 방법 등을 표시하는 화면

[출처: 스튜어트 러셀, 피터 노빅.(2016). 인공지능: 현대적 접근방식(제3판)]

2 지능 에이전트 체험하기

생활 속에서 체험할 수 있는 지능 에이전트를 찾아보고, 지능 에이전트와 협업해 보자.

1 생활 속에서 경험해 본 지능 에이전트를 찾아 빙고판을 완성해 보자.

구글이나 네이버, 다음 등의 검색 엔진을 사용한 적이 있다.	지문이나 얼굴 인식으로 스마트 기기를 잠금 해제한 경험이 있다.	목적지를 찾아가기 위해 지도앱을 이용한 적이 있다.	앱을 이용해 노래 인식을 해 본 적이 있다.	인공지능 챗봇과 대화해 본 적이 있다.
페이스북에서 자동으로 제품을 추천 받아본 적이 있다.(받았다면 그 이유는 무엇일까요?)	“중요” 라벨이 붙은 이메일을 받은 적이 있다.	카카오톡에서 추천 광고를 본 적이 있다.	컴퓨터(인터넷)에 글을 작성할 때 글자 자동 완성 또는 자동 교정을 해 본 경험이 있다.	음악 스트리밍 사이트에서 자동 추천된 음악을 들어본 경험이 있다.
온라인 쇼핑몰에서 A상품을 구매한 뒤, B상품을 추천받은 적이 있다.	수신한 메일이 스팸 메일로 분류된 적이 있다.(실제로 스팸 메일이었나요?)		SNS 광고를 클릭한 적이 있다.(주로 어떤 광고가 보이는지 친구와 비교해 봅시다.)	뉴스앱에서 추천한 뉴스를 본 적이 있다.(주로 어떤 뉴스를 추천받는지 친구와 비교해 봅시다.)
구글의 안전 검색(Safe search) 서비스를 사용한 적이 있다.	지메일에서 이메일을 보낼 때 추천 답변을 본 적이 있다.	사진 필터앱을 사용해 본 적이 있다.(제일 좋아하는 필터는 무엇인가요?)	동작 감지 비디오 게임을 해 본 적이 있다.	지메일에서 답메일을 보낼 때 ‘중요 메일(Nudge)’ 자동 올리기 기능을 사용한 적이 있다.
구글이나 네이버, 다음 등의 검색 엔진에서 검색어 자동 완성하기를 사용한 적이 있다.	컴퓨터가 자동으로 평가하는 숙제를 해 본 적이 있다.	웹사이트나 앱에서 일기 예보를 확인한 적이 있다.	음성으로 문자 메시지를 보낸 경험이 있다.	문자를 그림으로 바꿔 주는 서비스를 이용해 본 적이 있다.

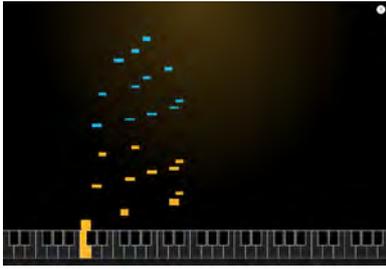
[출처: Blakeley H. Payne(MIT)(2019). 김한성 외 역, 모두를 위한 인공지능 윤리. KERIS]

도움 자료

- 인공지능 비서: 스마트폰에 탑재되어 있는 구글 어시스턴트나 애플 시리 등
- 인공지능 번역 프로그램: 국내에서 개발한 파파고나 구글의 번역기 등
- 로봇 기자: 국내에서 개발한 나리나 해외에서 개발한 워드 스미스 등
- 재난 구조 로봇: 국내에서 개발한 휴보, 해외에서 개발한 아르테미스 등
- 인공지능 의사: 국내에서 개발 중인 닥터 앤서, 해외에서 개발한 닥터 왓슨 등

2 지능 에이전트와 협업하면서 작동 원리에 대해 생각해 보자.

A.I. Duet



- ① 검색 엔진에서 AI duet을 검색한 뒤, 사이트에 접속한다.
- ②  를 클릭한다.
- ③ 키보드의 키를 누르거나, 마우스로 건반을 클릭하여 피아노를 연주한다.
- ④ 파란색은 내가 연주한 것이고, 노란색은 지능 에이전트가 연주한 것이다.

⑤ 지능 에이전트가 연주하는 원리가 무엇인지 생각해 보자.

Auto Draw



- ① 검색 엔진에서 Auto Draw를 검색한 뒤, 사이트에 접속한다.
- ②  을 클릭한다.
- ③ 하얀 도화지 영역에 그림을 그린다.
- ④ 화면 상단에 지능 에이전트가 추천하는 그림 중 원하는 그림을 선택한다. 
- ⑤ 색칠도 하고 글자도 넣어서 그림을 완성한다.
- ⑥ 메뉴를 눌러서 이미지를 다운받는다.

⑦ 지능 에이전트가 그림을 추천하는 원리가 무엇인지 생각해 보자.

3 지능 에이전트의 특징을 환경, 인식, 결과의 관점으로 정리하여 써 보자.

지능 에이전트	어떤 환경인가?	무엇을 인식할까?	나타나는 결과는?

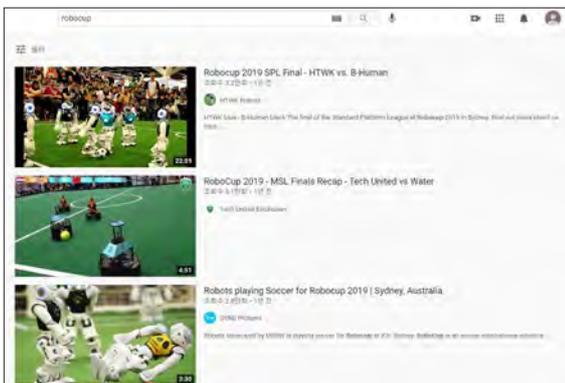
3 지능 에이전트 분석하기

지능 에이전트를 구상하고 분석해 보자.

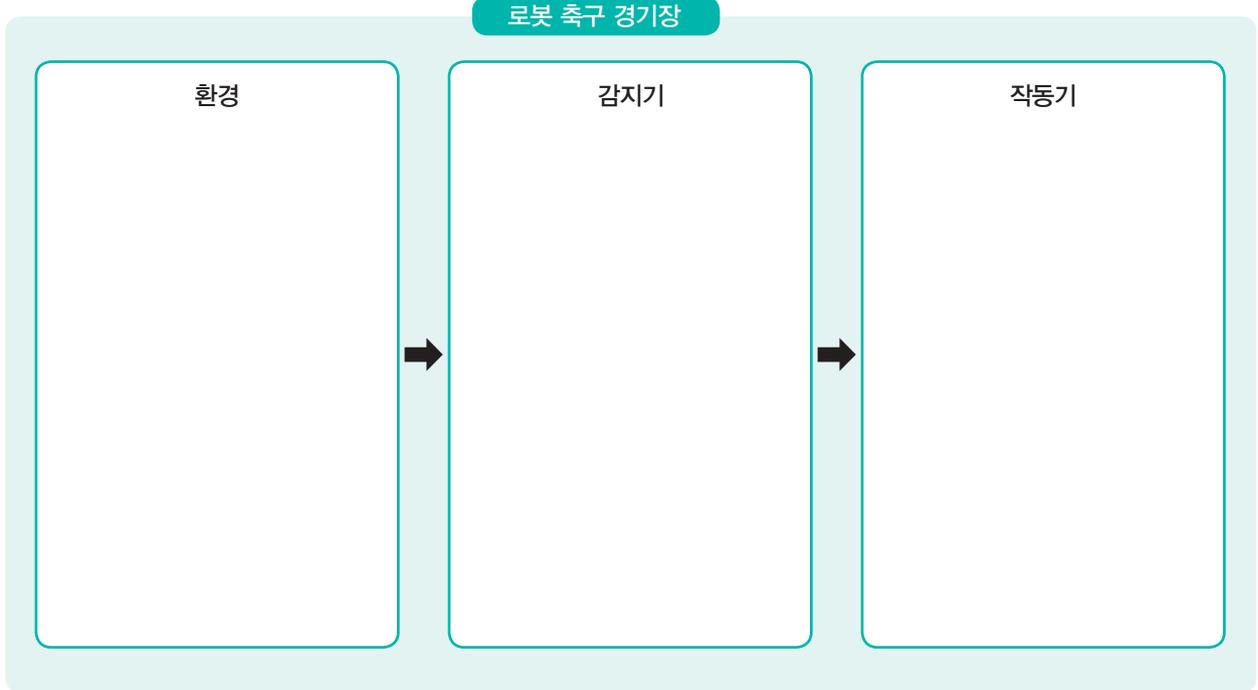
1 인간의 인식과 행동의 결과를 분석한 뒤, 이와 비교하여 지능 에이전트들을 분석해 보자.

<p>인식해야 할 것</p>	<p>축구 선수의 환경</p>  <ul style="list-style-type: none"> • • • • 	<p>행동의 결과로 나타나는 것</p>
<p>인식해야 할 것</p>	<p>인공지능 축구 로봇의 환경</p>  <ul style="list-style-type: none"> • • • • 	<p>행동의 결과로 나타나는 것</p>

2 유튜브에서 Robocup 2020을 검색하여 관련 영상을 찾아보고, 그 특징을 정리하여 써 보자.



3 인공지능 축구 로봇을 환경, 감지기, 작동기의 관점으로 좀 더 구체적으로 분석해 보자.



이솜 (ebssw.kr)에 접속한 뒤,
 배움터 - 이솜 스페셜 - 생활 속에서 찾아보는
 인공지능 이야기에서 응용 사례를 살펴보고
 다음과 같이 정리해 보세요.



I 지능 에이전트

4 다른 에이전트는 어떤 환경에서 어떤 역할을 하는지 탐색해 보자.

에이전트	환경	감지기	작동기

4 지능 에이전트의 사례 찾기

다양한 지능 에이전트의 사례를 찾아 체험해 보자.

1 대화형 AI 체험하기

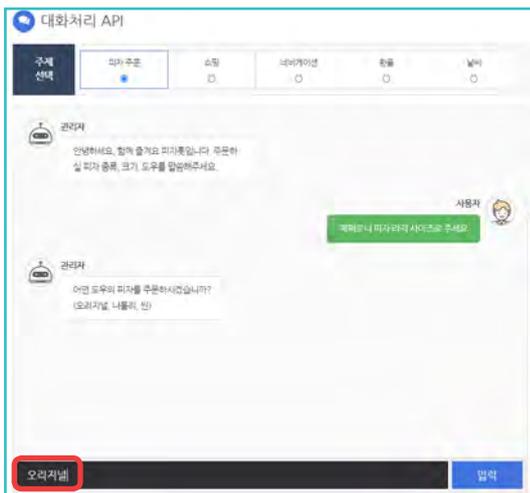
- 1 공공 인공지능 허브 사이트(<https://www.aihub.or.kr>)에 접속한다.
- 2 AI SW - 전체메뉴보기 - 개발가이드 - 대화처리를 클릭한다.



- 3 데모를 클릭하여 대화처리 API에 접속한다.



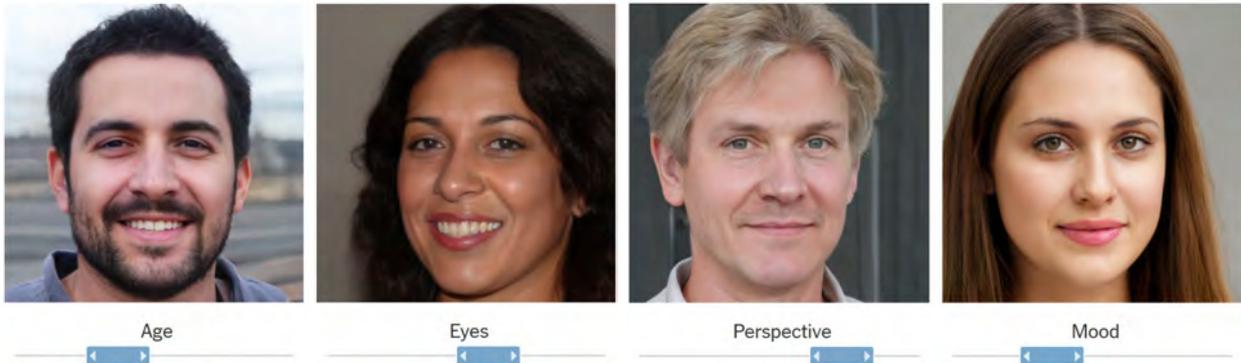
- 4 주제 선택에서 피자 주문 챗봇을 선택한 후, 관리자의 질문에 따라 피자를 주문해 본다.



- 5 챗봇이 어떤 원리로 작동하는지 추측해 보자.

2 이미지 생성 에이전트

- 1 검색 엔진에서 fake face를 검색한 뒤, 뉴욕 타임즈에서 공개한 fake faces 사이트에 접속한다.
- 2 페이지 중간쯤으로 이동해서 특성을 변경할 수 있는 부분을 찾는다.



- 3 나이, 눈 모양, 시점, 기분, 성별, 인종과 민족 항목의 스크롤을 좌우로 조절하면서 인공지능이 어떻게 사람의 얼굴을 변형하는지 살펴보자.
- 4 이미지 생성 에이전트가 만든 사람 얼굴이 실제와 얼마나 비슷한지 비교해 보자.
- 5 이미지 생성 에이전트가 사람의 얼굴을 만들어 내는 원리를 추측해 보자.

3 무인 탐사 에이전트(화성 탐사선)

- 1 검색 엔진에서 experience curiosity를 검색한 뒤, 사이트에 접속한다.
- 2 왼쪽 메뉴에서 Free Drive and Rover Replay를 눌러서 탐사선이 어떻게 움직이는지 관찰해 보자.
- 3 지능 에이전트 중 로봇이나 자율주행자동차가 작동하는 원리를 추측해 보자.



5 지능 에이전트 역할 보고서 작성하기

지금까지 조사하고 체험해 본 지능 에이전트를 분석하여 역할 보고서를 작성해 보자.

지능 에이전트 이름

지능 에이전트의 목적

지능 에이전트의 환경

지능 에이전트의 감지기, 작동기

지능 에이전트의 역할

6 나에게 필요한 지능 에이전트 구상하기

작성한 지능 에이전트 역할 보고서에 따라 지능 에이전트를 구상해 보자.

1 모둠을 구성한 후, 모둠별로 디자인 사고 과정을 통해 나에게 필요한 지능 에이전트를 구상해 보자.



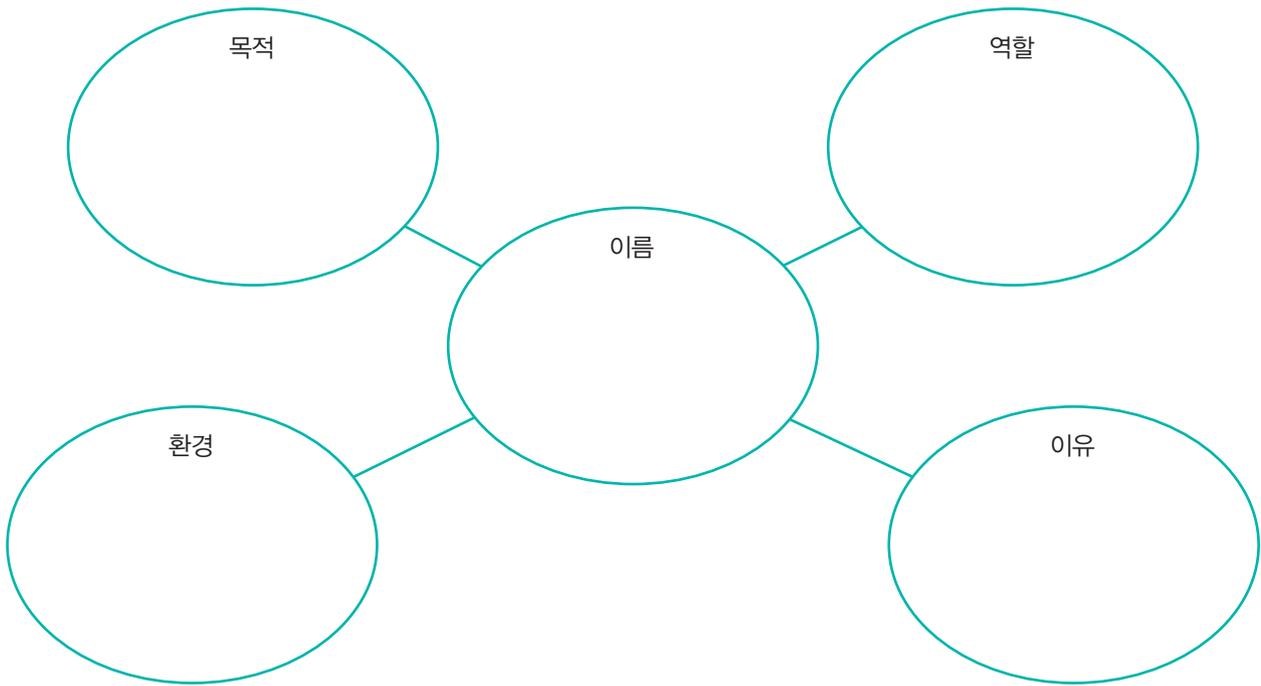
2 모둠원끼리 자신에게 지능 에이전트가 필요한 상황과 원하는 결과를 공유하고 써 보자.

모둠원 이름	지능 에이전트가 필요한 상황	원하는 결과

3 모둠에서 설계할 지능 에이전트 중 한 가지를 선정하고, 그 이유를 써 보자.

지능 에이전트 이름	이유

4 모듈에서 선택한 지능 에이전트를 정의해 보자.



5 각자 설계 아이디어를 스케치해 본 후, 친구들에게 설명해 보자.

7 발표 및 정리하기

1 모둠별로 설계한 지능 에이전트를 발표하고 서로 평가해 보자.

※ 평가도에 따라 별표에 색칠해 보세요.

모둠 이름	지능 에이전트 이름	창의성	활용성	실행 가능성
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆
		☆☆☆☆☆	☆☆☆☆☆	☆☆☆☆☆

※ 창의성: 지능 에이전트가 창의적인지 평가한다.

※ 활용성: 지능 에이전트의 목적이 명확하고, 활용 가능성이 높은지 평가한다.

※ 실행 가능성: 지능 에이전트를 실제로 구현할 수 있는 가능성이 높은지 평가한다.

2 지능 에이전트를 설계하는 활동을 통해 알게 된 점을 써 보자.



3단계 공감 나누기



성찰하기

지능 에이전트 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

1 배운 내용을 마인드맵으로 표현하고 친구들과 비교해 보자.

2 프로젝트를 진행하면서 알게 된 지능 에이전트의 장점과 단점을 써 보자.

지능 에이전트의 장점

지능 에이전트의 단점

3 프로젝트를 진행하면서 더 알고 싶은 점이 있다면 그 내용을 써 보자.

어떤 마음가짐으로 참여하였으며, 또 제대로 이해하며 활동하였는지 서로의 생각을 쓰도록 하세요.



알파고는 인공지능 프로그램을 개발하는 딥마인드(DeepMind)에서 개발한 바둑을 두는 인공지능 에이전트이다. 2016년 대한민국 서울에서 이세돌 사범과의 대국을 통해 세상에 알려지게 되었다.

인공지능의 역사를 보면 인간과 인공지능은 몇 차례 대전을 펼쳤는데, 1997년 체스 경기에서 인간 대표인 카스파로프와 IBM에서 만든 인공지능 딥 블루(Deep Blue)가 대결하여 인간이 패배하였다. 2011년에는 미국의 퀴즈쇼에서 IBM의 왓슨(Watson)이 인간 챔피언들을 물리치고 우승하였다. 2016년에는 알파고가 바둑 대결에서 인간을 4:1로 이긴 성과를 보여 주었다.

바둑은 인류의 문화유산 중에 역사가 깊은 경기로 높은 수준의 창의력과 집중력을 필요로 하는 경기이고, 바둑의 경우의 수는 대략 2×10^{170} 으로 인공지능이 연산을 통해서 해결할 수 없을 것이라고 생각했다. 쉽게 이해할 수 있도록 비교해 보면 우주 전체의 원자 개수가 대략 12×10^{78} 라고 하는데, 바둑의 경우의 수는 이보다 훨씬 큰 수이기 때문에 인공지능이라고 하더라도 쉽게 계산할 수 없을 것이라고 생각했던 것이다. 하지만 전 세계 2억 명의 사람들이 지켜보는 가운데 알파고가 승리하면서 인공지능의 발전상을 보여 주었다.



[출처: 딥마인드 <https://deepmind.com/alphago-korea>]

알파고는 2017년에 바둑 챔피언인 중국의 커제와 겨뤄서 이긴 후 은퇴한다. 이후 딥마인드에서는 알파고 제로(AlphaGo Zero)를 만들었는데, 알파고 제로는 알파고와 다른 방식으로 학습해서 72시간 만에 기존의 알파고를 물리치는 성과를 보여 준다. 기존의 알파고는 인간의 바둑 데이터 16만 건을 학습하는 데 수개월이 걸렸지만, 알파고 제로는 초기 데이터 없이 72시간을 스스로 학습한 후, 기존 알파고와의 대국에서 100전 100승을 거두었다. 알파고 제로는 바둑뿐만 아니라 체스, 쇼기(일본식 장기 게임)도 할 수 있다. 2020년에 발표된 뮤제로(MuZero)는 계획 기능을 갖춘 지능 에이전트로 발전하였다. 뮤제로는 바둑, 체스, 쇼기뿐만 아니라 아타리 게임에도 도전하여 좋은 성과를 올리고 있다.

알파고의 발전이 우리에게 어떤 영향을 주게 될지 살펴볼 필요가 있다.

II 인식

1 단계 핵심 개념 알기

- 센서와 인식
- 컴퓨터 비전
- 음성 인식
- 언어 이해



2 단계 창의융합적 사고로 해결하기

- 활동 1 센서는 어떻게 주변 환경을 인식할까?
- 활동 2 인공지능은 어떻게 이미지를 인식할까?
- 활동 3 인공지능은 어떻게 음성을 인식할까?
- 활동 4 인공지능은 어떻게 사람의 말을 이해할까?

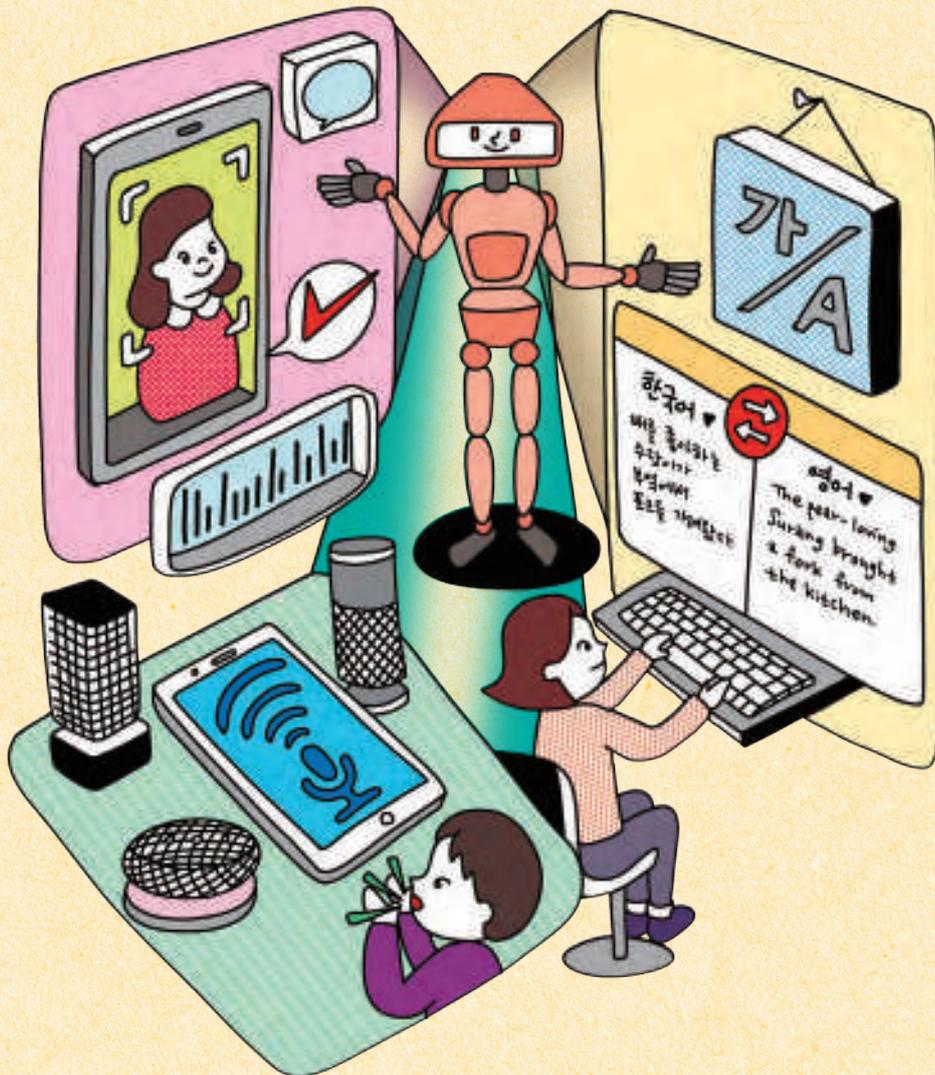
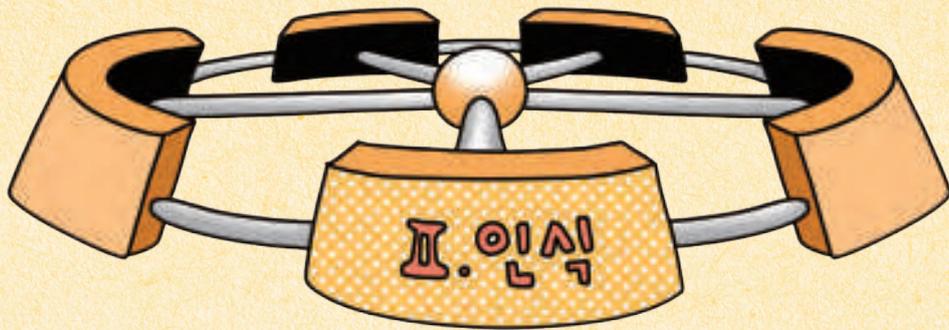


3 단계 공감 나누기

- 비주얼 씹킹으로 정리하기

👁️ 인식 단원에서는 무엇을 배울까요?

다양한 센서의 동작과 이미지 인식, 소리 인식, 언어 처리 등을 체험하며, 인공지능이 환경과 상호작용하는 방법을 배웁니다.





1단계 핵심 개념 알기

인식

✔ 인공지능은 어떻게 세상과 소통할까?

지능 에이전트는 센서로 환경을 인식하여 세상과 상호작용한다.

1 센서와 인식

Q 개념 탐색 센서는 어떻게 주변 환경을 인식할까?

센서는 자연계에 존재하는 물리량 또는 대상물의 정보를 측정해 전기 신호로 바꾸는 장치이다. 센서에 처리 기능을 추가하여 인식하고 상태를 판정하는 것을 감지라고 한다.

스마트폰 센서의 종류에는 동작, 위치, 환경 센서 등이 있다.

유형	기능
동작	동작 센서는 사용자 움직임을 감지하고, 가속도와 중력, 회전 속도, 회전 벡터 값, 편향 등을 측정한다.
위치	위치 센서는 기기의 물리적인 위치를 확인하는 데 사용된다. 지자기 센서와 가속도계를 활용해서 나침반 기능을 제공하고, 기기 위치를 확인한다.
환경	환경 센서에는 습도, 조도, 압력, 온도 등을 알려 주는 4개의 센서가 있다. 온도계는 기기 내부의 온도를 측정하여, 스마트폰이 과열되는 위험을 방지한다.

2 컴퓨터 비전

Q 개념 탐색 인공지능은 어떻게 세상을 볼까?

컴퓨터 비전(Computer Vision)은 인간의 시각처럼 컴퓨터 기기가 자율적으로 할 수 있도록 만든 분야이며, 인간과 다른 시각 기능을 부여하거나 각기 다른 능력으로 새로운 시각을 만들고, 인간의 한계를 보조하는 기술이다. 이로서 자율주행뿐만 아니라 컴퓨터의 눈이 필요한 여러 분야에서 활용된다.



“지능 에이전트는 센서, 시각, 청각 등의 인식을 통하여 세상과 상호작용한다.”

3 음성 인식

Q 개념 탐색 인공지능은 어떻게 소리를 인식할까?

음성 인식(Speech Recognition)이란 사람이 말하는 언어를 컴퓨터가 해석해 그 내용을 문자로 변환하는 것을 말한다.



▲ 인공지능 스피커의 음성 인식 과정

음성 인식 기술은 전처리와 인식의 두 가지 단계로 구성된다. 전처리는 사용자가 발성한 음성으로부터 인식할 대상이 되는 구간을 찾아 잡음을 제거하고, 소리의 특징을 추출하는 과정이다. 인식은 음성에서 특징을 추출하여 패턴을 분석하고 언어를 처리하는 과정을 포함한다.

4 언어 이해

Q 개념 탐색 인공지능은 어떻게 언어를 이해하고 처리할까?

인공지능이 사람의 말을 이해하고, 처리할 수 있는 능력의 기반은 자연어 처리(Natural Language Processing) 기술이다.

자연어 처리는 기계인 컴퓨터가 인간의 언어를 이해할 수 있도록 하는 것으로, 자연어 이해와 자연어 생성의 두 가지로 구성된다. 자연어 이해는 의미적 역할과 같은 콘텐츠에서 메타 데이터를 추출해 기계가 인간의 언어를 이해하고 분석하는 데 도움을 준다. 자연어 생성은 컴퓨터화된 데이터를 자연어 형태로 변환하는 번역기 역할을 한다.

이러한 자연어 처리 기술이 적용된 대표적인 사례로는 기계 번역, 챗봇, 스팸 감지, 감정 분석, 워드 클라우드, 맞춤법 수정 등이 있다.

활동 미리 보기

핵심 개념과 관련된 활동은 어떤 내용으로 전개될까요?

활동 1

센서는 어떻게 주변 환경을 인식할까?

- 다양한 스마트폰 센서 활용하기
- 다양한 센서 활용 아이디어 표현하기

활동 2

인공지능은 어떻게 이미지를 인식할까?

- 티처블 머신에서 사물 인식 체험하기
- 이미지 인식 앱으로 사물을 인식하기
- TensorSpace.js로 손글씨 인식하기
- 컴퓨터 비전의 한계

활동 3

인공지능은 어떻게 음성을 인식할까?

- 티처블 머신에서 사람 목소리 구분하기
- 음성 인식으로 글쓰기
- 인공지능 스피커와 대화하기

활동 4

인공지능은 어떻게 사람의 말을 이해할까?

- 번역기 체험하기
- 워드 클라우드로 말의 주요 의미 파악하기



2단계 창의융합적 사고로 해결하기

활동 1 센서는 어떻게 주변 환경을 인식할까?

- ▶ 다양한 센서를 활용하여 센서가 어떻게 주변 환경을 인식하는지 체험해 보고, 센서의 활용과 한계를 설명할 수 있다.

1 문제 이해하기

다음 기사를 읽고 스마트 시티에서 센서가 어떻게 활용되고 있는지 알아보자.

최근 자율주행자동차에 사용되는 라이다(LiDAR) 센서와 레이더(Ladar) 센서가 스마트 시티에 다양하게 적용되고 있다. 도로의 교통량 측정, 횡단보도에서 보행자 보호, 지하철 스크린 도어 개폐 제어 등에 활용된다.



▲ 도시의 안전을 위해 사용되는 자율주행 센서

사례 1 도로에서 보행자를 감지하는 라이다 센서

가끔 보행자가 걸어가면서 스마트폰을 보다가 신호등의 신호를 깜박 놓쳐 빨간 불일 때 횡단보도에 진입하는 경우가 있다. 이때 신호등의 보행 신호 음성 안내 장치에 설치된 라이다 센서는 이를 감지하고 보행자에게 경고를 알려 준다. 독일에서는 가로등에 카메라 센서와 라이다 센서를 설치해 차량이나 보행자 정보를 파악하고 있다. 미국에서는 신호등에 라이다 센서를 설치하여 교통 데이터를 분석함으로써 보행자와 운전자의 안전에 도움을 주고 있다.

사례 2 스마트 홈과 도시의 교통량 파악에 사용되는 레이더 센서

기존에는 도시에 CCTV 등 카메라가 많이 적용되어 사생활 침해 문제가 있었는데, 라이다와 레이더 센서를 활용하면 카메라로 발생할 수 있는 사생활 침해 문제를 해결하면서도 효율적으로 교통량을 파악할 수 있다는 장점이 있다. 라이다 센서는 해상도가 높지만, 상대적으로 날씨 조건에 약하다. 레이더는 라이다에 비해 해상도가 떨어지지만, 날씨 조건에 상대적으로 강하고 인체에 미치는 영향이 거의 없다.

[출처: LG CNS 블로그, <https://blog.lgcns.com/2400>]

이외에 스마트 시티에 활용될 수 있는 센서에는 무엇이 있는지 생각해 보자.

무엇을 알아야 할까?

배경 지식

센서와 인식

인공지능은 주변 환경과 상호작용하기 위해서 센서를 이용한다. 센서(Sensor)는 자연계에 존재하는 물리량 또는 온도, 압력, 습도, 거리 등 대상물의 정보를 측정해 전기 신호로 바꾸는 장치로, 센서에서 처리 기능을 추가하여 인식하고 상태를 판정하는 것을 감지(Perceive)라고 한다.

자율주행자동차는 카메라와 다양한 센서 등을 활용하여 주변의 환경과 사물을 인식한다. 카메라를 통해 다른 자동차와 표지판, 교통 신호를 인식하고, 레이더가 주변 사물을 빠르게 인식해 갑작스러운 상황에 대처할 수 있다.



▲ 자율주행자동차에 내장된 다양한 센서

	<p>라이다 센서 빛을 발사한 뒤 물체에 부딪혀 되돌아오는 빛의 속도를 측정해 물체와의 거리를 측정한다.</p>
	<p>레이더 센서 전자파를 발사한 뒤 물체에 부딪혀 돌아오는 반사파로 물체와의 거리, 속도, 방향 등을 파악한다.</p>
	<p>카메라 라이더 센서, 레이더를 통해 알 수 없는 정확한 영상 정보를 얻는다.</p>
	<p>GPS 차량의 위치 정보를 얻는다.</p>

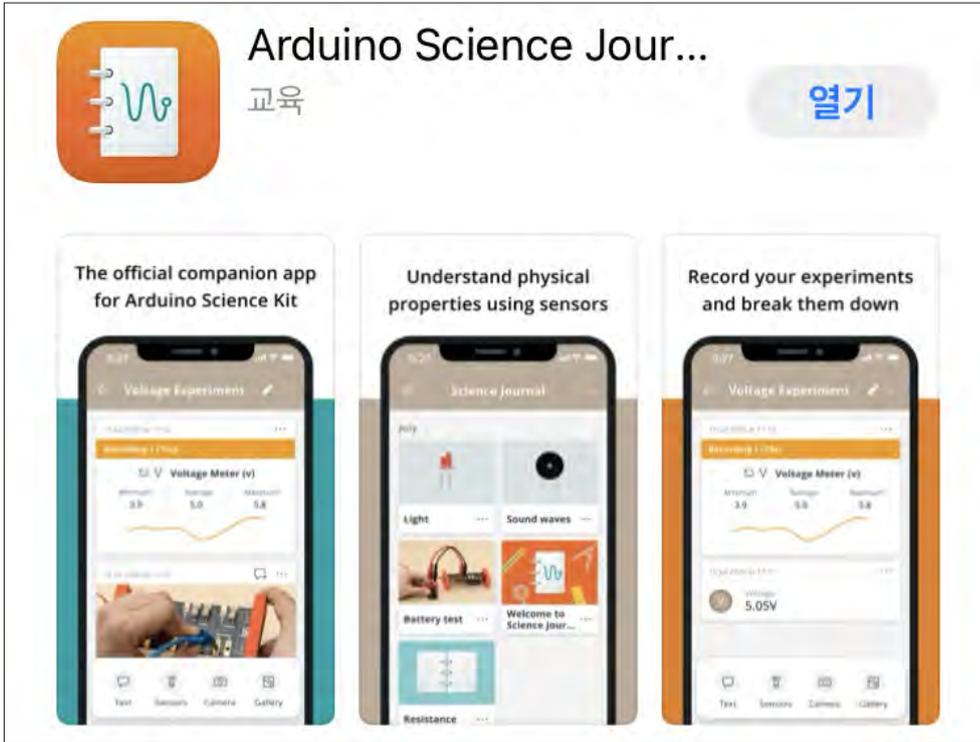
▲ 자율주행자동차에 내장된 센서의 기능

2 스마트폰 센서 활용하기

스마트폰에는 주변 환경을 인식할 수 있는 다양한 센서가 내장되어 있다. 과학 저널 (Science Journal) 앱을 설치해서 스마트폰의 다양한 센서를 체험해 보자.

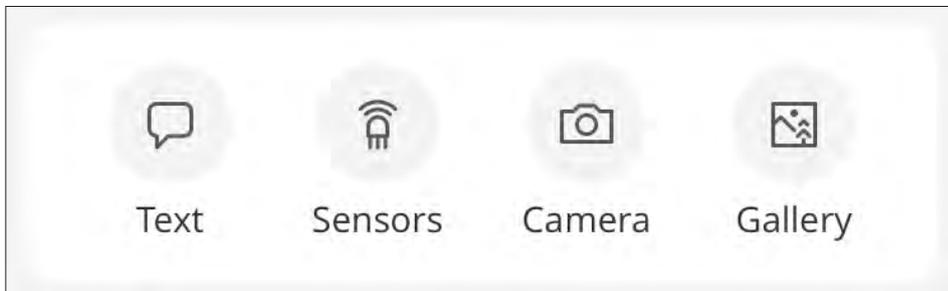
1 과학 저널 앱을 설치한 후, 어떠한 기능이 있는지 살펴보자.

① 앱스토어에서 'Arduino Science Journal'을 검색하여 앱을 설치한다.



▲ 앱 설치하기

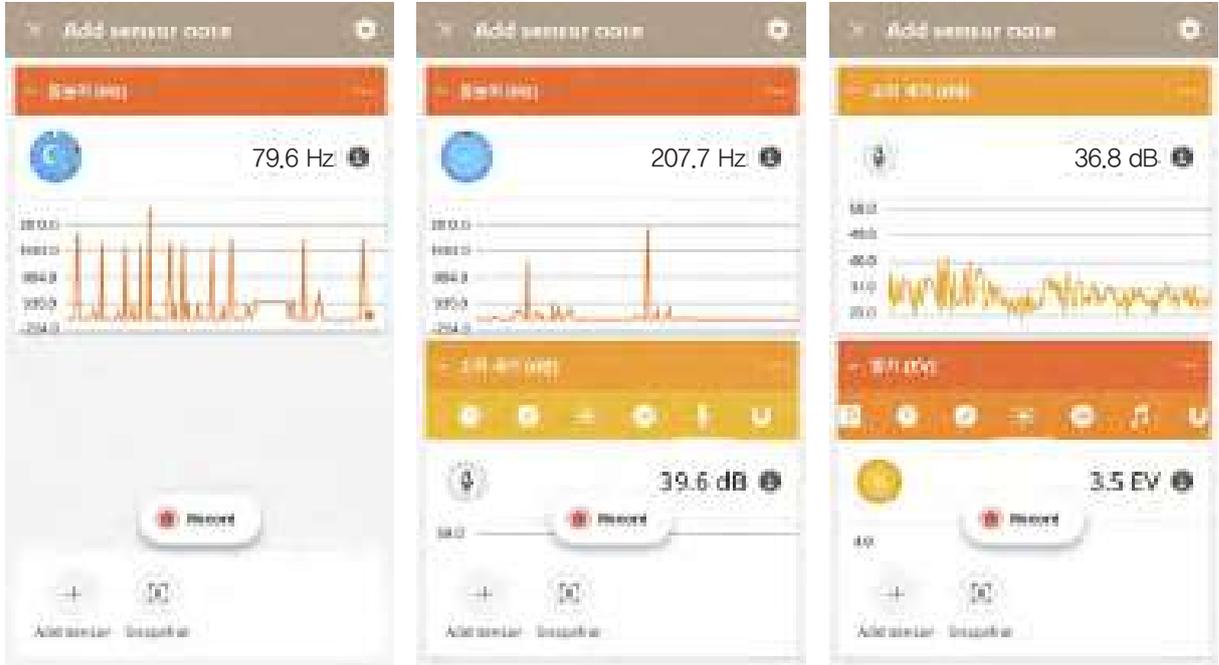
② 과학 저널 앱을 이용하면 스마트폰에 내장된 가속도, 기압계, 나침반, 밝기, 소리 세기, 자기장을 측정할 수 있는 다양한 센서로 과학 실험 데이터를 얻을 수 있다.



▲ 과학 저널 앱의 다양한 기능

2 과학 저널 앱에서 스마트폰에 내장된 센서들이 주변 환경에 대한 정보를 어떻게 인식하는지 비교해 보자.

① 친구들과 대화하거나 음악을 재생하여 같은 소리가 어떻게 다른 값으로 측정되는지 비교해 보자. 여러 센서가 측정한 값을 동시에 그래프로 비교하여 나타낸다.



▲ 소리 음높이 측정하기

▲ 소리 음높이와 소리 세기 비교

▲ 소리 세기와 빛의 밝기

② 같은 소리로 음높이와 세기 값이 어떻게 다른지 기록하고, 측정한 값이 무엇을 의미하는지 적어 보자.

센서	측정값	센서로 측정한 값의 의미
음높이(Pitch)	207.7Hz	
소리 세기(dB)	39.6dB	

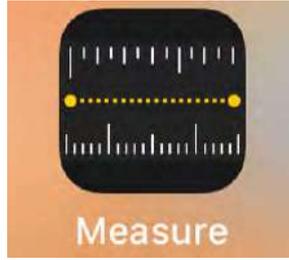
소리 센서로 측정한 값을 기록하여 들어보세요.



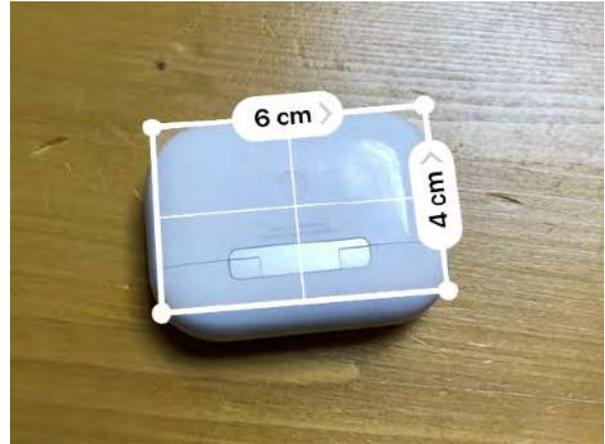
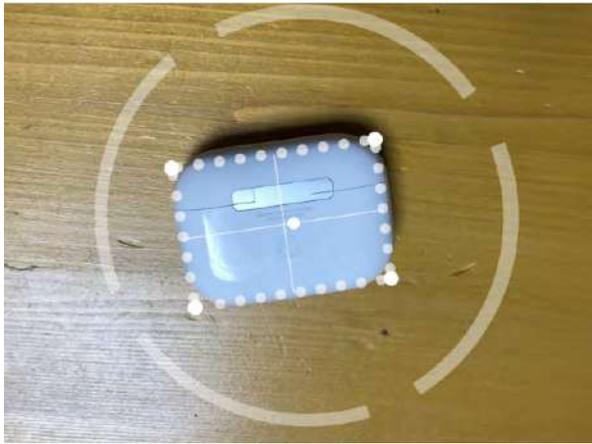
③ 과학 저널 앱에서 밝기(Brightness)를 측정한 값이 어떻게 달라지는지 기록하고, 측정한 값이 무엇을 의미하는지 적어 보자.

센서	측정값	센서로 측정한 값의 의미
어두울 때	2.4	
밝을 때	1.1	

3 스마트폰에 내장되어 있는 측정(Measure) 앱을 이용하여 사물의 크기를 인식해 보자.



1 그림과 같이 1개의 물체를 인식하여 물체의 가로세로 길이를 측정해 보자.



2 2개의 물체를 동시에 인식하여 측정하는 실험을 해 보자. 연필처럼 길고 가는 것은 자동으로 인식하지 못하는 경우에는 **+**를 눌러서 물체의 끝과 끝에 포인터를 추가하여 길이를 측정한다.



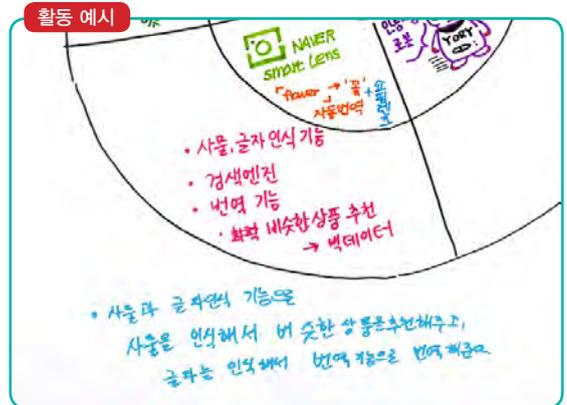
4 과학 저널과 측정 앱에서 센서가 동작하는 방식의 차이는 무엇인가? 어느 것에 오류가 많은지 토의해 보자.

과학 저널 앱	측정 앱

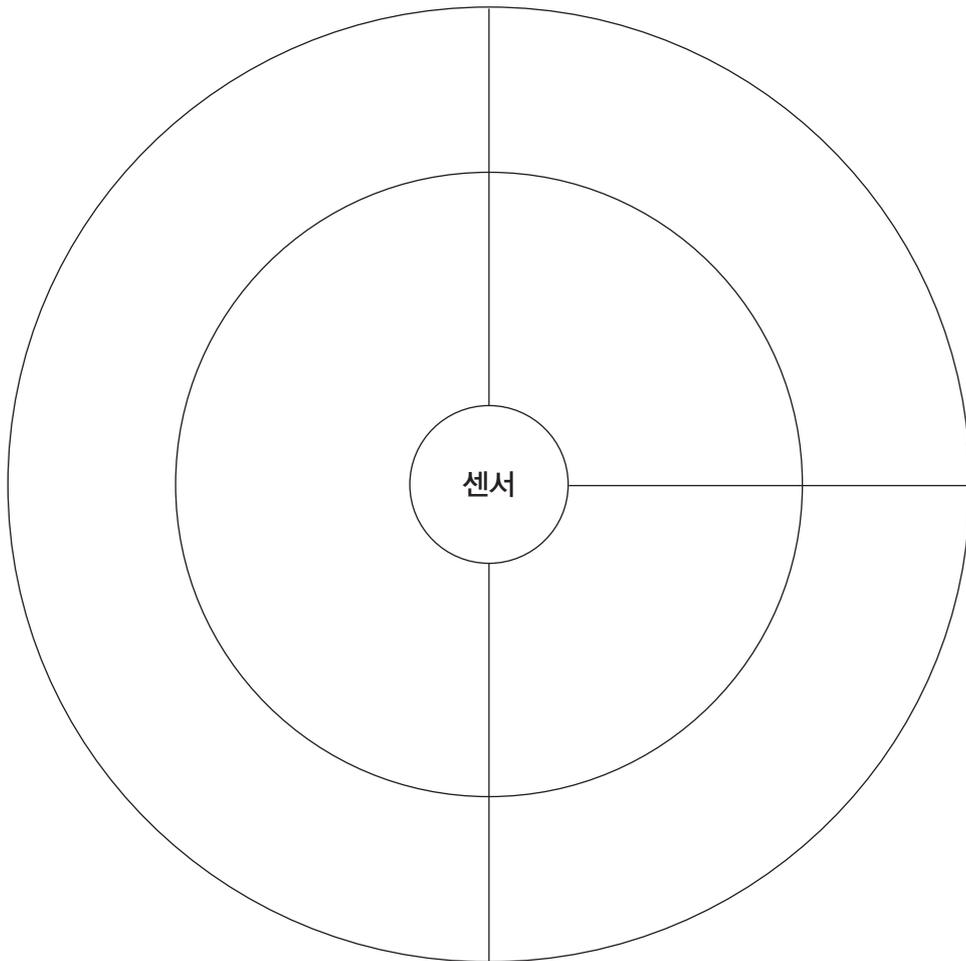
3 센서 활용 아이디어 표현하기

인공지능이 센서를 활용할 수 있는 사례와 아이디어를 비주얼 씹킹(Visual Thinking)으로 표현해 보자.

- ① 센서 활용에 관한 주제를 작성한다.
- ② 센서 활용 모습을 글과 그림으로 표현한다.
- ③ 어떤 센서가 활용될 수 있는지, 어떤 정보를 감지할 수 있는지 작성한다.
- ④ 인공지능 활용 아이디어를 글로 표현한다.



• 주제: _____



4 정리하기

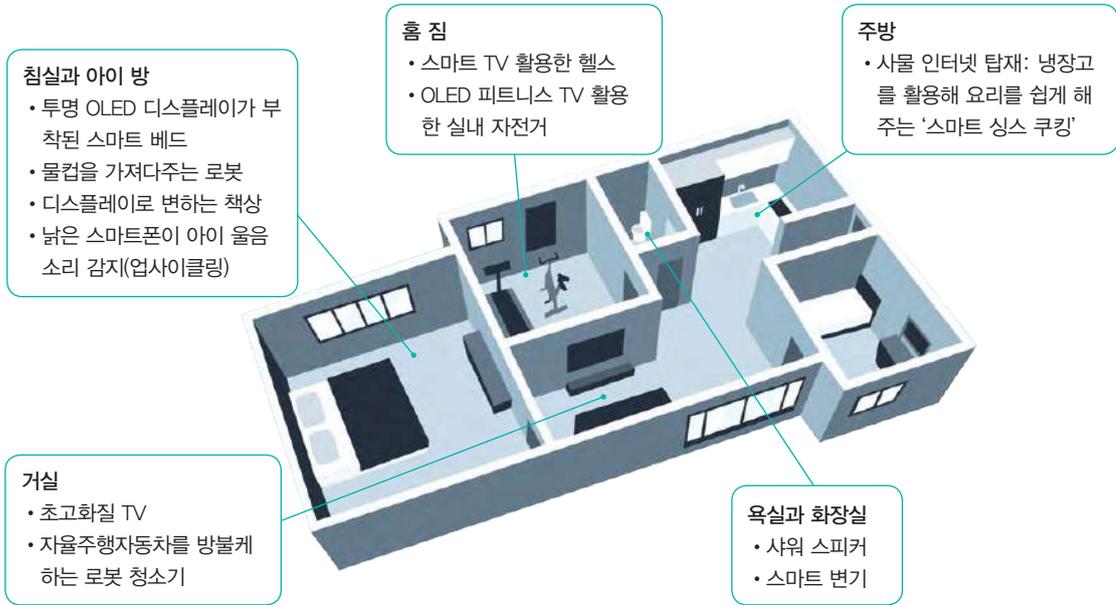
1 이 활동을 통해 알게 된 점을 써 보자.

- 센서는 _____ 이다.
- 스마트폰 센서의 유형에는 (), (), () 등이 있으며, () 정보를 인식하여 경로를 탐색하고, () 센서를 이용하여 주변 환경의 소음을 측정할 수 있다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
준비	활동 전 배경 지식을 충분히 이해하였는가?					
센서와 인식	다양한 센서를 활용하여 주변 환경의 정보를 측정해 보았는가?					
센서 활용	인공지능의 센서 활용 아이디어를 설명할 수 있는가?					

음료를 가져다주는 로봇, 침대 발끝에서 올라오는 투명 디스플레이, 자율주행자동차에 적용된 센서가 장착돼 스스로 움직이는 로봇 청소기... 세계 최대 국제전자제품박람회인 'CES 2021'에서 공개된 미래의 '똑똑한 집'을 들여다보자.



이른 아침 알람 소리에 눈이 떠졌다. 알람이 울린 곳은 침대 발끝에 위치한 투명 디스플레이. 머리맡에 있는 버튼을 눌러 알람을 끄자 디스플레이에는 기온과 습도 등 날씨 정보가 나타났다. 버튼을 다시 누르자 간밤에 뒤척이진 않았는지 수면 패턴을 담은 정보가 디스플레이 화면에 나타났다. 잠들기 전 보던 영화를 마저 봐야겠다고 생각하고 리모컨 버튼을 누르자 디스플레이가 더 높이 올라와 55인치 TV로 바뀌었다.

영화를 보다 보니 목이 말랐다. “물 마시고 싶어.”라고 외치자 관절이 있는 긴 팔과 집게손을 가진 로봇이 시원한 물을 가져다주었다. 주방에서 침실까지 로봇이 이동하는 길 한가운데에는 벗어둔 옷가지가 장애물처럼 놓여 있었지만 로봇은 이리저리 피해 겁을 날랐다. 이후 집게손으로 옷가지를 집어 세탁기에 집어넣었다.

어느덧 밥 먹을 시간. 스마트폰을 들고 주방으로 갔다. 스마트폰으로 쿠킹 어플리케이션을 열자 추천 식단과 상세한 조리법이 떴다. 조리법을 골라 요리를 시작하면 가열 시간, 온도 등이 오븐 등 조리기기로 바로 전송된다. 부족한 재료는 냉장고 스크린으로 곧바로 주문할 수 있다. 음식 사진을 스마트폰으로 촬영하자 함께 마시기 좋은 추천 음료 목록이 나타났다. 음식을 만든 뒤 “테이블 세팅해 줘.”라고 말하자 로봇이 수저와 식기를 테이블에 놓았다. 식사를 다했다고 말하자 로봇이 정리까지 도왔다.

[출처:동아일보. 2021. 1. 16 기사]

활동 2

인공지능은 어떻게 이미지를 인식할까?

- ▶ 이미지 인식과 컴퓨터 비전의 원리를 이해하고, 이미지 인식의 한계에 대해 설명할 수 있다.

1 문제 이해하기

사례 두 가지를 읽고, 인공지능 이미지 인식의 원리와 한계를 생각해 보자.

사례 1



과일 농가에서 과일을 등급별로 분류하는 일에도 인공지능이 적용되고 있다. 딥러닝과 컴퓨터 비전을 활용한 ‘농산물 자동 선별기’는 농산물의 종류와 품질을 인식해 자동으로 과일과 야채를 분류해 준다. 농산물의 모양과 색깔, 크기에 따라 다양한 등급으로 선별하는 것은 물론, 농산물에 생긴 흠집과 멍 등을 감지해 상품성이 떨어지는 것을 자동으로 골라낸다. 이처럼 카메라에 상품을 인식하고 자동으로 선별하는 과정에서 제품 손상을 줄일 수 있다.

사례 2

고속도로에서 자율주행 모드로 직진하던 자율주행자동차가 반대편 차선에서 좌회전하는 대형 트레일러 트럭을 인지하지 못해 교통사고가 일어났다. 자율주행자동차는 하얀색 대형 트레일러 트럭의 옆면 아랫부분과 정면 충돌하였는데, 그 이유는 당시 날씨가 매우 화창해 자율주행 제어 장치나 운전자 모두 하얀색 트레일러 트럭을 차량으로 인식하지 못해 사고가 난 것으로 보고 있다.



[출처: <https://news.joins.com/article/20251575>]

과일 분류기로 어떻게 과일을 분류할 수 있을까? 그리고 이러한 기술의 한계는 무엇인지 생각해 보자.

무엇을 알아야 할까?

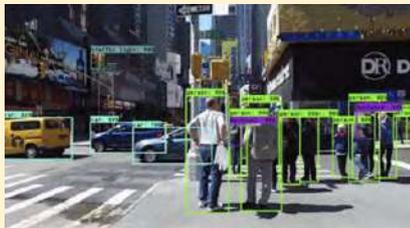
배경 지식

컴퓨터 비전과 이미지 인식

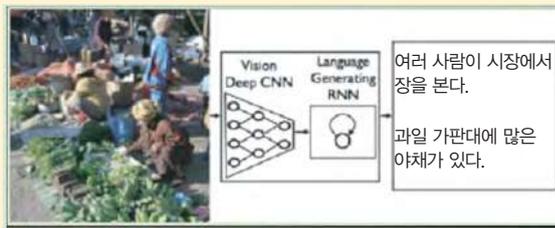
컴퓨터 비전이란 이미지나 영상과 같은 시각 데이터를 컴퓨터가 입력받아 이해하도록 처리하는 분야이다. 인간의 시각이 할 수 있는 일을 보조하거나 대체하는 자율적인 시스템을 만드는 것을 목표로 한다. 컴퓨터 비전은 글씨와 같은 2차원 흑백 패턴 인식부터 2차원 도형이나 사진을 분류하는 것, 3차원 공간에서 물체를 인식하고 추적하는 것 등 다양하다. 컴퓨터 비전은 무인 마트, 자율주행자동차, 로봇, 생활 속 카메라와 스마트 기기 등에 활용된다.

이미지 인식은 이미지의 정보를 식별하는 기술로, 컴퓨터 비전의 핵심 기술 중 하나이다. 이미지에 포함된 객체를 인식하고 분류하는 것이 대표적이다.

컴퓨터 비전에서 이미지 캡셔닝(Image Captioning)이란 컴퓨터가 사진을 보고 적절한 설명을 자동으로 붙이는 기술이다. 캡셔닝 알고리즘은 이미지를 입력하면 이미지를 설명하는 문장을 출력한다.



▲ 자율주행자동차와 컴퓨터 비전



▲ 이미지로 이야기 생성

인공지능은 아래 왼쪽 사진에 ‘서핑 보드를 들고 있는 남성’, 오른쪽 그림에는 ‘들판 속 보리 확대 사진’이라는 제목을 붙였다. 인공지능 기술의 발달로 다양한 이미지를 인식 및 분석해 콘텐츠를 자연어로 해설해 주는 것이 가능해졌다.



[출처: 인공지능 신문, <https://www.aitime.kr/news/articleView.html?idxno=18091>]

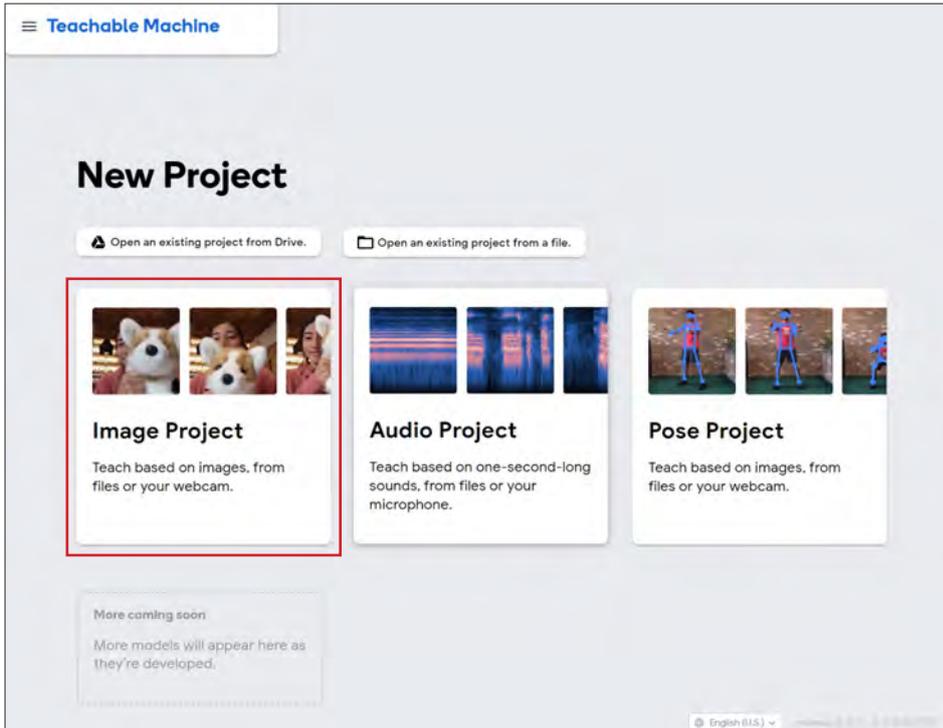
이미지 인식에는 합성곱 신경망(CNN, Convolution Neural Network)이라는 알고리즘이 대표적으로 활용되고 있으며, 영상 인식에 특화된 다층 신경망이다. 어린 아이가 개와 고양이의 모습을 보고 배우는 것처럼 인공지능도 많은 데이터를 통해 개와 고양이의 특성을 추출하여 단계적으로 학습한다.

2 사과일까? 귤일까?

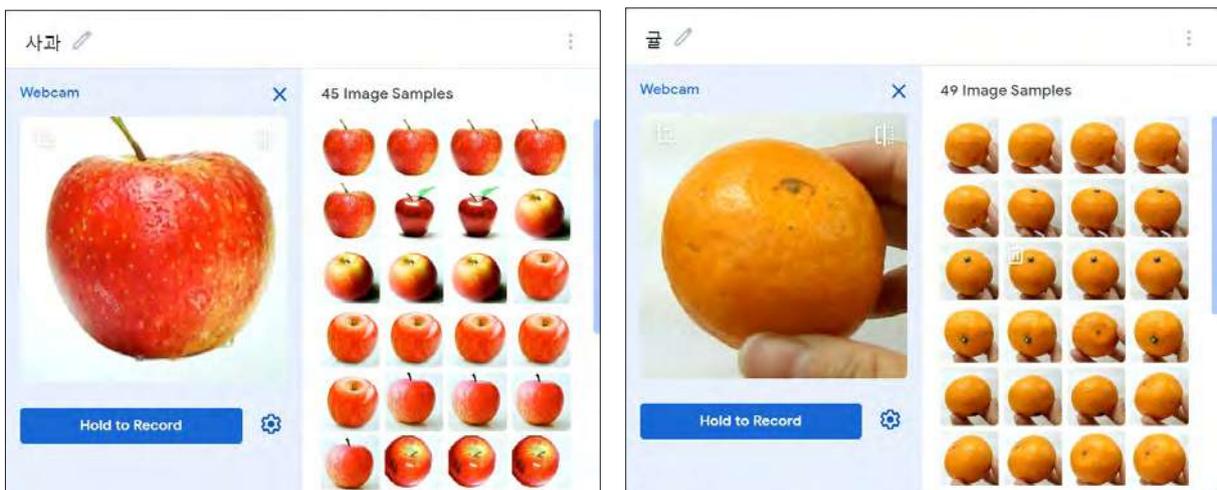
티처블 머신(<https://teachablemachine.withgoogle.com/>)의 이미지 프로젝트를 이용하여 인공지능이 사물을 어떻게 인식하고 분류하는지 실습해 보자.

1 데이터 준비하기

- 1 티처블 머신에서 Image Project를 선택한다.



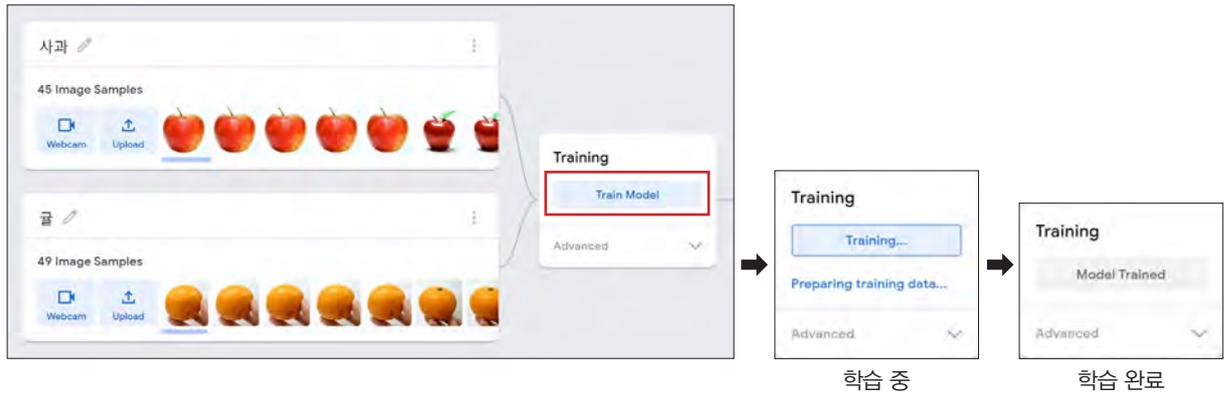
- 2 클래스1에는 '사과', 클래스2에는 '귤'을 입력한 뒤, [Webcam] - [Hold to Record]를 클릭하여 훈련 이미지의 샘플을 추출한다. 웹캠으로 직접 이미지 데이터의 샘플을 얻는다.



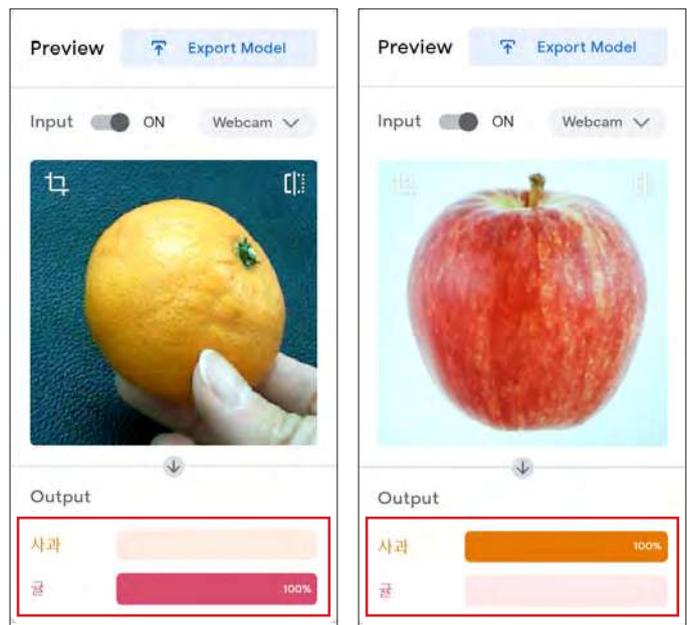
직접 사물을 촬영하여 훈련 이미지 샘플을 추출하기 어려운 경우에는 미리 수집한 이미지를 이용할 수 있다. 학습할 데이터는 많을수록 좋으며, 학습할 때는 두 객체의 이미지 샘플 수를 비슷하게 하여 추출한다.

2 이미지 학습과 테스트

- ① 화면 중간에 있는 [Train Model]을 눌러 사과와 귤 사진을 모델에 학습시키면, 모델이 사과와 귤을 구분할 수 있게 된다.



- ② 화면 오른쪽의 [Input(입력)] - [Webcam]을 클릭한 뒤, 학습할 때와 다른 사과와 귤 이미지로 모델을 테스트한다. 테스트 데이터를 입력하면 화면 오른쪽 아래에서 인식 결과를 곧바로 확인할 수 있다.



- ③ 학습하지 않은 전혀 다른 종류의 사물 이미지로 테스트하여 인식 결과를 확인해 보자.

- ④ 다른 종류의 사물을 잘 인식하였는가? 인식하지 못했다면 그 이유는 무엇일까?

이유	
----	--

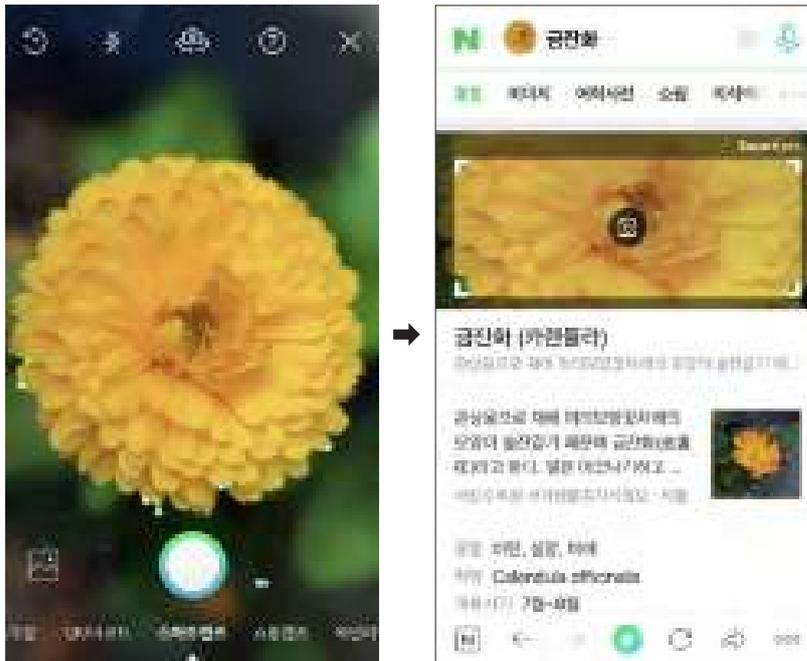


3 이미지 인식은 어디에 활용될까?

스마트폰 앱에서 물체를 인식하면 어떠한 정보를 알려 주는지 확인해 보자.

1 스마트 렌즈로 식물 이름 찾기

[네이버 앱] - [스마트 렌즈]를 실행한 뒤, 식물 사진을 찍으면 사진 속 식물이 무엇인지 알려 준다.



다양한 물체를 인식해 보고, 어떤 결과가 나오는지 확인해 보자.

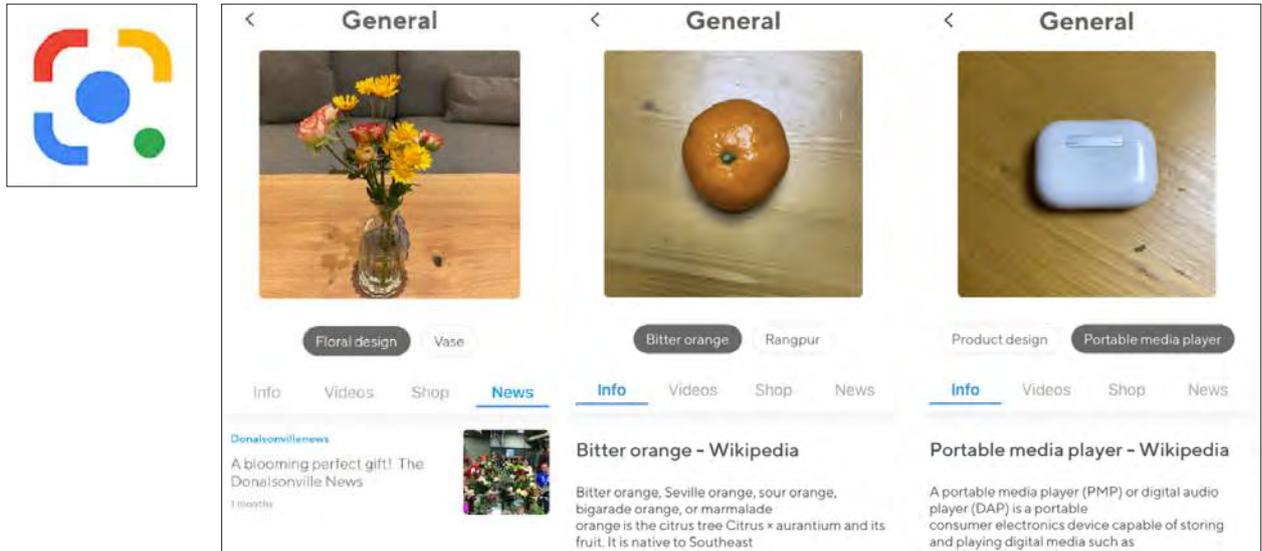
인식한 물체	결과
	예 줄리아 페페 화분을 인식시켰더니 줄리아 페페 식물에 대한 정보를 제공하고, 연관 식물도 추천하였다.
	

카메라로 촬영한 사진 속 사물을 인식할 수 있는지 실험해 보세요



2 구글 렌즈로 다양한 사물 인식하기

구글 렌즈(Google Lens)는 사용자가 보고 있는 내용을 이해하고 해당 정보를 이용하여 텍스트를 복사 또는 번역할 수 있다. 또 식물과 동물을 식별하고, 시각적으로 유사한 이미지를 찾아 준다. 구글 렌즈를 활용하여 사물을 인식해 보자.



인식한 물체	결과
	<p>예 줄리아 페페 화분을 인식시켰더니 줄기 식물(stem plant)로 인식하고, 식물의 종류를 정확히 구분하지는 못했다.</p>
인공지능은 어떻게 이미지를 인식할까?	

3 구글 렌즈와 스마트 렌즈를 활용하여 여러 가지 사물을 인식해 보고, 어떤 차이가 있는지 기록해 보자.

	스마트 렌즈	구글 렌즈
인식이 잘 되는 경우		
인식이 잘 안 되는 경우		
이용할 수 있는 기능		

4 만약 인식이 잘 안 된다면 그 이유가 무엇인지 친구와 이야기해 보자.

4 이미지 인식은 어떤 과정으로 이루어질까?

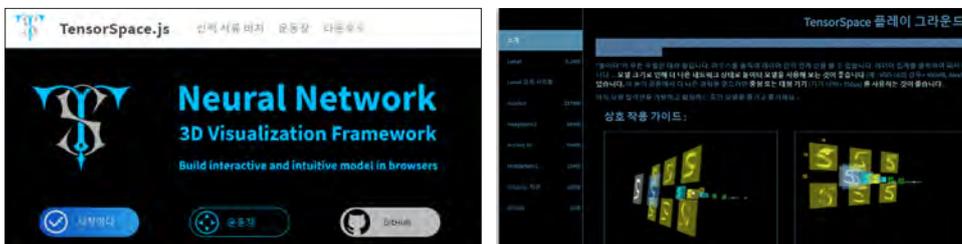
이미지 인식은 딥러닝을 사용하는데, TensorSpace.js를 이용하면 다양한 딥러닝 모델을 이용하여 숫자나 그림을 인식하는 과정을 3차원으로 볼 수 있다.

TensorSpace.js는 프로젝트 모델에서 어떤 식으로 학습이 되는지를 보여 주는 라이브러리이며, 네트워크 층을 3차원으로 시각화하여 신경망에서 딥러닝의 추상적인 인식 과정을 구체적으로 확인할 수 있다.

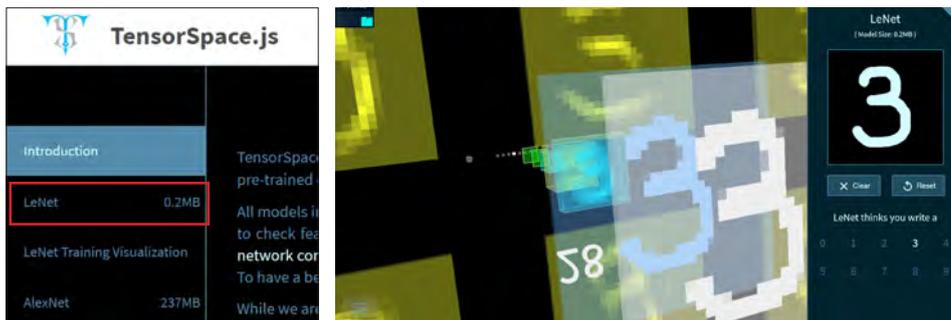
TensorSpace의 Playground는 사전 제작된 다양한 모델을 제공하여 학습이 이루어지는 과정을 시각적으로 확인할 수 있는 공간이다. 'Playground'에서 객체 분류, 객체 감지, 이미지 생성 등을 포함한 다양한 사전 훈련된 딥러닝 모델을 경험할 수 있다.

1 TensorSpace.js에 접속하여 손글씨 데이터 입력하기

- 1 크롬(Chrome)이나 엣지(Edge)에서 <https://tensorspace.org/>에 접속한 뒤, Playground를 클릭한다.



- 2 왼쪽 메뉴에서 'LeNet'을 선택하면 화면이 바뀌면서 오른쪽에 숫자를 쓸 수 있는 창이 나타난다.



'LeNet'은 CNN개념을 최초로 제안한 Yann LeCun이 개발한 모델이다. LeNet 모델을 사용하여 인식할 숫자를 오른쪽 입력창에 그리면 신경망을 통해 숫자를 인식하고 사용자가 어떤 숫자를 그렸는지 예측한다. 또한 각 레이어마다 어떤 과정을 거쳐 신경망이 동작하는지도 관찰할 수 있다.

2 인식 결과 확인하기

- 1 손글씨 숫자를 1에서 9까지 인식해 보고, 인식이 잘되는 경우와 인식이 어려운 경우는 무엇인지 비교해 보자.

인식이 잘되는 경우	인식이 어려운 경우

- 2 인식이 어려운 경우, 그 이유가 무엇인지 생각해 보자.

5 정리하기

1 이 활동을 통해 알게 된 점을 작성해 보자.

- 이미지 인식은 _____이다.
- 이미지 인식을 체험할 수 있는 방법 _____한다.
- 스마트 렌즈와 구글 렌즈는 _____한다.
- TensorSpace.js를 이용하면 사전 훈련된 여러 모델을 이용하여 _____있다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
준비	활동 전 배경 지식을 충분히 이해하였는가?					
이미지 인식	다양한 방법으로 이미지 인식을 체험하였는가?					
이미지 인식의 한계	이미지 인식의 한계를 설명할 수 있는가?					

활동 3

인공지능은 어떻게 음성을 인식할까?

▶ 음성 인식 프로그램을 이용하여 음성 인식의 과정과 원리를 설명할 수 있다.

1 문제 이해하기

다음 대화를 읽고, 문제가 발생한 이유를 알아 보자.



인공지능 스피커를 사용하면 같은 질문에도 다르게 대답하는 경우가 많은데, 왜 다르게 대답하는지 그 이유를 생각해 보자.

.....

.....

무엇을 알아야 할까?

배경 지식

음성 인식

음성 인식(Speech Recognition)이란 사람이 말하는 음성 언어를 컴퓨터가 해석해 그 내용을 문자 데이터로 전환하여 처리하는 기술로, STT(Speech-to-Text)라고도 한다.

우리 생활 속에서 음성 인식은 다양한 곳에 활용된다. 자동차 내비게이션의 목적지를 설정하거나, 스마트 TV로 채널을 변경하거나 스마트폰에서 연락처를 찾고, 인공지능 스피커와 대화할 때에도 음성 인식이 활용된다.

음성 인식은 크게 전처리와 인식이라는 두 가지 단계로 구성된다. 인식은 세부적으로 특징 추출, 패턴 분석, 언어 처리의 과정을 거친다.

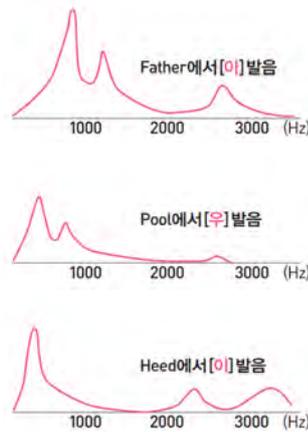
전처리

음성 전처리

다양한 소리가 섞여 있을 때 사람의 귀는 원하는 소리만 집중해서 들을 수 있지만, 기계는 중요한 소리뿐 아니라 주변에서 들리는 잡다한 소리를 모두 저장한다. 전처리에서 사용자가 발성한 음성으로부터 인식 대상이 되는 구간을 찾아 잡음을 제거한다.

특징 추출

인식 과정을 위해 소리의 특징을 추출한다. 기계가 음성을 인식하기 위해서는 소리 주파수를 분석해야 한다.



인식

패턴 분석

분석을 통해 특징을 추출하면 사람마다 목소리와 발성, 억양 등이 다르므로 다양한 환경에서 통계적인 정보를 추출하여 기준이 되는 패턴을 생성한다.

언어 처리

미리 학습한 음성 모델 데이터베이스와 유사도를 검사해서 가장 적합한 것을 문자 또는 명령어로 변환한다.

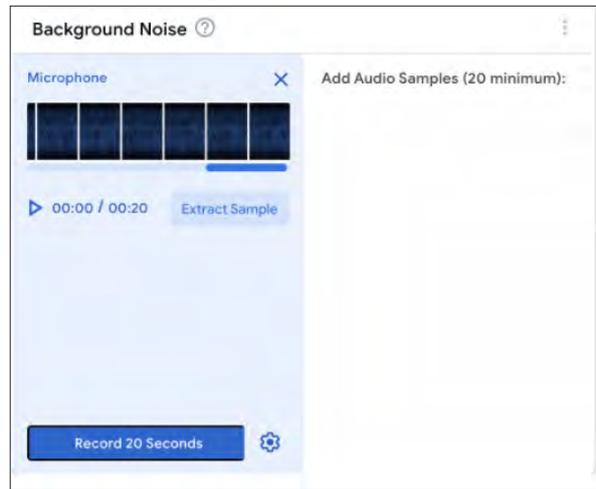
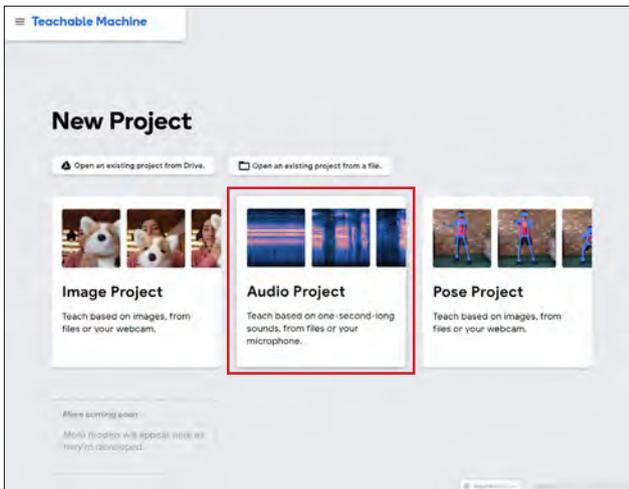
2 인공지능이 사람의 목소리를 구분할 수 있을까?

티처블 머신(<https://teachablemachine.withgoogle.com>)의 오디오 프로젝트를 이용하여 두 사람의 목소리를 구분할 수 있는지 알아보자.

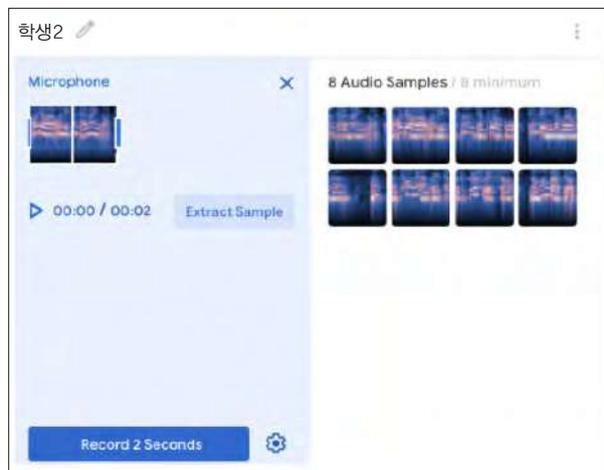
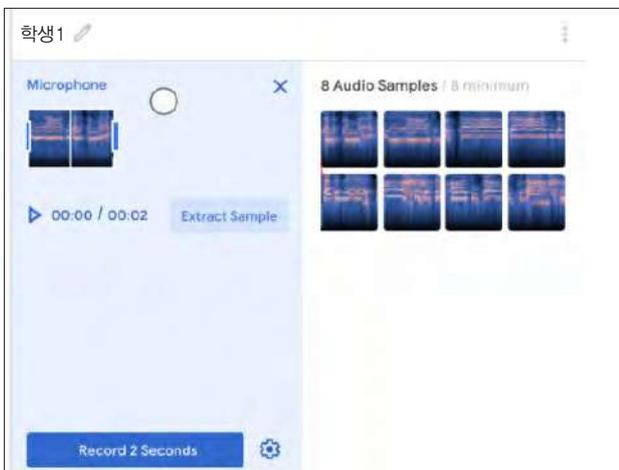


1 데이터 준비하기

- ① 티처블 머신에서 Audio Project를 선택한다.
- ② 사람이 소리를 내지 않을 때에는 배경 잡음으로 분류하기 위해 20초 동안 배경 잡음 샘플을 추출한다.

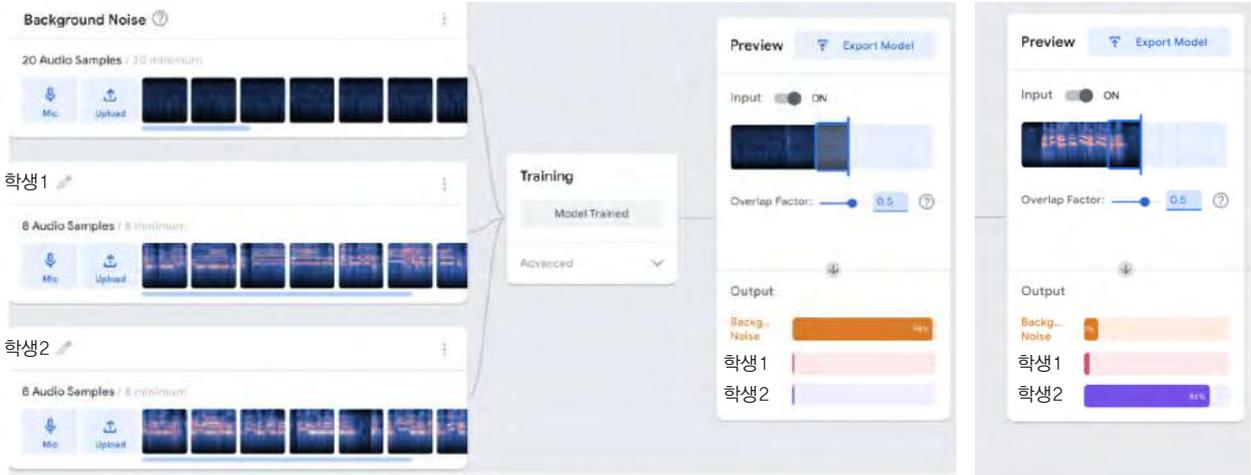


- ③ '학생1' 클래스를 만들고, **Record 2 Seconds**를 눌러 학생1의 목소리를 2초 동안 녹음을 한다. **Extract Sample**을 눌러 학습을 위한 최소 8개의 샘플을 추출한다.
- ④ 마찬가지로 방법으로 '학생2' 클래스를 만들고, 학생2의 목소리로 녹음을 한다. 학습을 위해 최소 8개의 샘플을 추출한다.



2 소리 학습시키기

- ① 화면 중간에 있는 [Train Model]을 눌러 배경 잡음과 학생 두 명의 목소리를 학습시킨다.
- ② 학습이 끝나면 바로 소리를 인식하기 시작한다. 학생1과 학생2의 목소리를 바꾸어서 테스트한다.



▲ 배경 잡음과 학생의 목소리

- ③ 나의 목소리와 친구의 목소리로 평상시 대화할 때 목소리, 기쁜 목소리와 화난 목소리를 학습시켜 보고, 인공지능이 잘 구분하는지 확인해 보자.

	대화할 때 목소리	기쁜 목소리	화난 목소리
나의 음성			
친구의 음성			

- ④ 사람의 음성 인식은 인공지능의 음성 인식과 어떻게 다른지 작성해 보자.

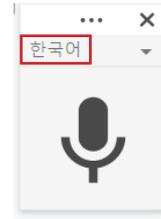
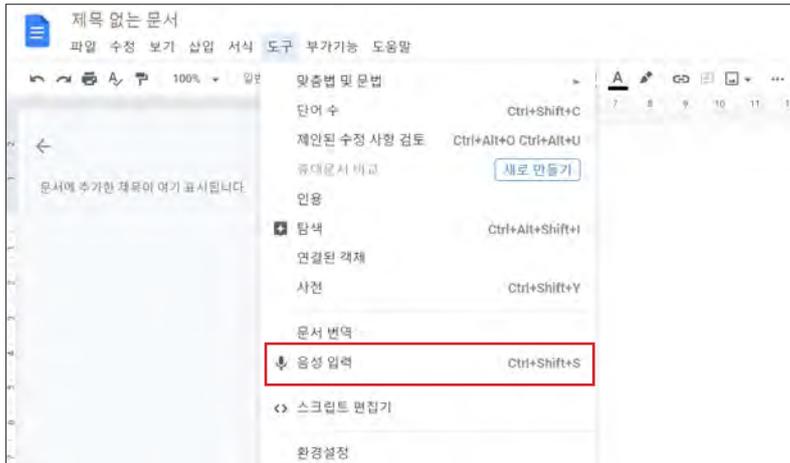
사람의 음성 인식	인공지능의 음성 인식

3 음성 입력으로 글쓰기

스마트폰이나 마이크가 있는 컴퓨터를 이용하면 말로 문서를 작성할 수 있다. 구글 문서를 이용하여 음성으로 시를 입력해 보자.

1 한국어 음성 입력하기

- 1 구글 문서를 실행하여 [도구]-[음성 입력]을 이용하면 다음과 같이 마이크 모양이 나타난다. 마이크를 눌러 말하는 것을 문서로 작성할 수 있다.



- 2 한국어를 선택한 상태에서 다음 시를 음성으로 입력해 보자.

별 헤는 밤

윤동주

계절이 지나가는 하늘에는
가을로 가득 차 있습니다.

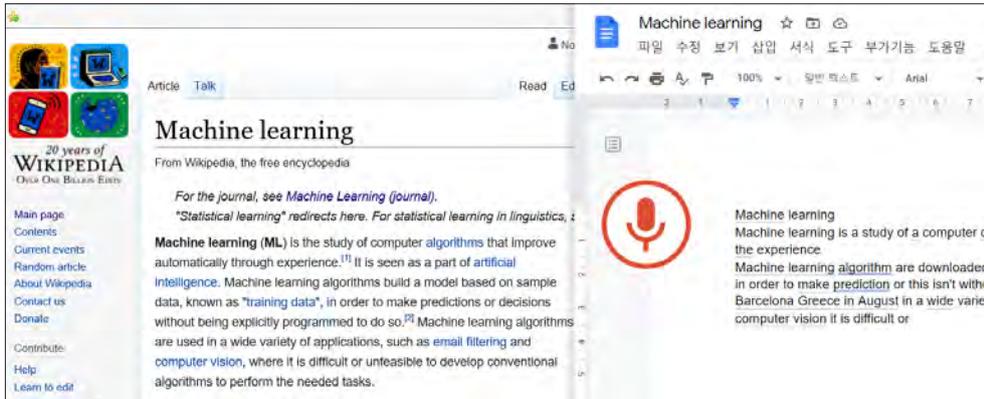
나는 아무 걱정도 없이
가을 속의 별들을 다 헤는 듯합니다.

가슴 속에 하나 둘 새겨지는 별을
이제 다 못 헤는 것은
쉬이 아침이 오는 까닭이요,
내일 밤이 남은 까닭이요,
아직 나의 청춘이 다하지 않은 까닭입니다.

별 하나에 추억과
별 하나에 사랑과
별 하나에 쓸쓸함과
별 하나에 동경과
별 하나에 시와
별 하나에 어머니, 어머니,
(중략)

2 영어 음성 입력하기

- 1 영어로 말하는 것을 구글 문서로 입력할 수 있는지 알아보자. 먼저 위키백과에서 'Machine Learning'이라는 용어를 검색하여 나타나는 영어 원문을 보고, 구글 문서에 음성 입력으로 문서를 작성해 보자.



- 2 영어로 말하는 것을 구글 문서에 잘 입력할 수 있었는지 친구와 비교해 보자. 만약 입력이 잘 안되었다면 그 이유가 무엇인지 생각해 보자.

	나	친구
목소리로 문서 작성이 잘되는가?		
입력이 잘되는 것은 어떤 경우인가?		
입력이 잘 안되는 것은 어떤 경우인가?		



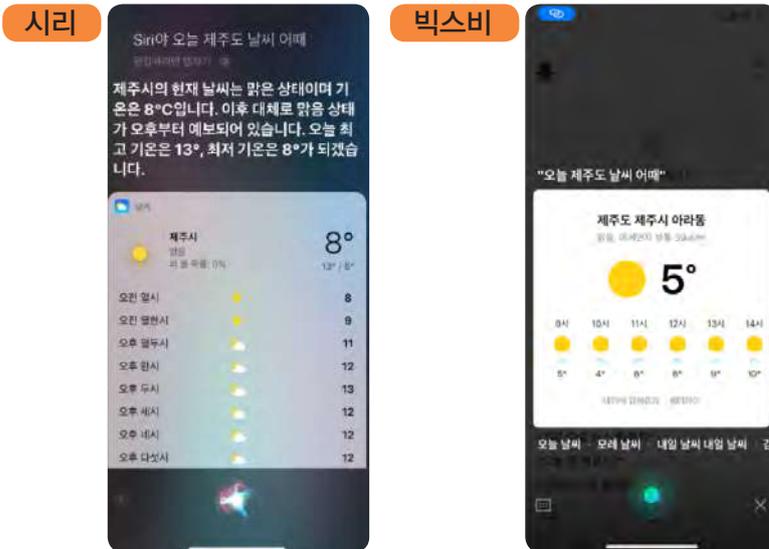
- 3 음성 입력할 때 좋은 점과 어려운 점을 담아 인공지능에게 편지를 작성해 보자.

4 인공지능 스피커와 대화하기

우리 생활 속에서 쉽게 접할 수 있는 인공지능 스피커에게 질문해 보고, 같은 질문에 어떻게 다르게 대답하는지 비교해 보자.

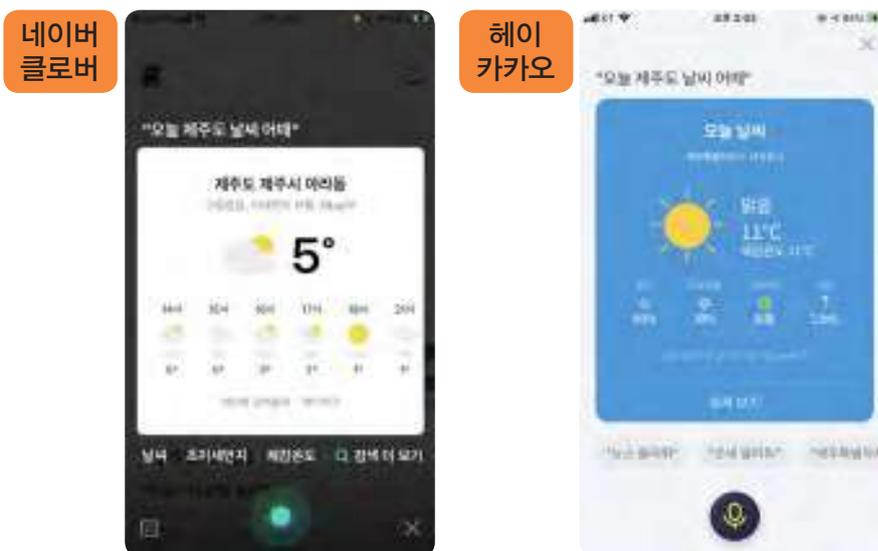
1 인공지능 스피커에게 질문하기

- ① 시리와 빅스비에게 “오늘 제주도 날씨 어때?”라고 질문하면 어떻게 대답하는지 비교해 보자.



시리	
빅스비	

- ② 네이버 클로버와 헤이 카카오 앱을 설치하여 “오늘 제주도 날씨 어때?”라고 질문하면 어떻게 대답하는지 비교해 보자.



네이버 클로버	
헤이 카카오	

네이버 클로버나 헤이 카카오는 별도의 스피커 기기를 통해 음성으로 정보를 알려 줍니다.



2 인공지능 스피커와 대화하기

인공지능 스피커에게 대답하기 어려운 질문을 해 보며, 대답하는 데 어떤 한계가 있는지 실험해 보자.

- ① 인공지능 스피커를 호출한 뒤, 인공지능이 대답하기 어려운 질문을 해 보자. 그리고 다음 질문으로 '랩 해줘'라고 요청해 보자.
- ② 인공지능 스피커가 어떻게 대답하는지 비교해 보고, 다르게 대답하는 이유가 무엇일지 생각해 보자.



▲ 인공지능 스피커와 대화하기

	인공지능 스피커	대답
1	시리	
2	빅스비	
3	클로버	
4	헤이 카카오	

3 모두가 협력하여 인공지능 스피커의 동작을 4컷 만화로 표현해 보자.

스피커가 인식을 잘하는 경우와, 어떤 한계가 있는지 포함하여 만화로 표현해 보세요.



5 정리하기

1 이 활동을 통해 알게 된 점을 작성해 보자.

- 음성 인식은 _____ 이다.
- 인공지능으로 음성 인식을 할 때 좋은 점은 _____ 이다.
- 인공지능의 음성 인식에는 _____ 한계가 있다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
준비	활동 전 배경 지식을 충분히 이해하였는가?					
음성 인식	다양한 방법으로 인공지능의 음성 인식을 체험하였는가?					
음성 인식의 한계	인공지능 음성 인식의 한계를 설명할 수 있는가?					

읽을거리

인간과 가장 비슷한 언어를 구사할 수 있는 인공지능, 인간처럼 대화하는 자연어 처리 모델 세 가지

자연어 처리 AI 모델로는 구글에서 개발한 트랜스포머(Transformer)와 버트(BERT), 그리고 GPT-3(Generative Pre-trained Transformer 3)가 꼽힌다. 자연어 처리 AI는 자연어를 이용해 사람과 대화를 하거나, 소설이나 신문기사 등의 문장을 만들어내는 AI를 뜻한다.

인공지능이 사람 수준으로 자연어를 이해하고 구사하는 데에는 한계가 있다. 자연어와 컴퓨터 언어에는 큰 차이가 있기 때문이다. 가장 큰 차이는 '문맥'의 이해 여부다. 문맥은 단어 또는 문장의 앞뒤 상황에 따라 나타나는 언어적인 맥락이다. 같은 단어나 문장이라도 문맥에 따라 그 의미가 다르게 사용될 수 있는 요소를 말한다. 뛰어난 인공지능이라 할지라도 인간처럼 문맥을 이해하기는 어려운 일이다.

트랜스포머는 특정 언어를 다른 언어로 번역하는 '기계 번역'에 특화되어 있다. 버트는 자연어를 이해하는 데 특히 뛰어난 성능을 보인다. GPT-3는 자연어 생성을 목적으로 개발돼 언어 구사 능력에 뛰어나다.

2020년 오픈 AI에서 개발한 GPT-3는 언어 입력으로 다양한 텍스트 생성 작업을 수행할 수 있는 딥러닝 언어 모델로, 인간처럼 이야기를 쓸 수 있다. GPT-3는 양방향으로 자연어를 분석하는 버트와 달리 한방향으로 분석하는 단방향 모델이다. 상대적으로 자연어를 이해하는 성능은 부족하지만 차례로 문장을 만들어 나갈 수 있어 자연어 생성에 적합하다. 미국 채프먼대 학생은 GPT-3가 작성한 시나리오를 바탕으로 단편 영화를 제작하기도 했고, GPT-3가 직접 쓴 글이 신문기사나 칼럼으로 게재되기도 했다.

[출처: 과학동아 2021년 2월호, 현존 최고의 자연어처리 인공지능 선발대회]



활동 4

인공지능은 어떻게 사람의 말을 이해할까?

- ▶ 인공지능의 자연어 처리 기능을 활용하여 어떻게 언어를 이해하는지, 또한 한계점이 무엇인지 설명할 수 있다.

1 문제 이해하기

다음 기사를 읽고, 기계 번역의 성과와 한계에 대해 생각해 보자.

4명의 번역가와 3개의 인공지능 번역기가 각각 지문을 번역한 결과를 비교하는 실험을 하였다.

주어·명사구	부사어·부사구	서술어·동사구
배를 좋아하는 수량이가	/ 부엌에서 /	포크를 가져왔다
Su-rang who likes a boat	from kitchen	brought a fork
Su-rang who likes pears	from kitchen	got a fork
⋮		⋮

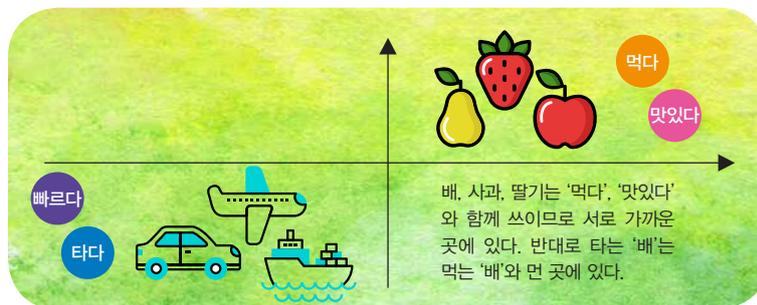
이 실험에서 전문 번역가는 30점 만점에 최고 25점을 받았지만, 인공지능은 15점을 넘지 못하였다.

인공 신경망 기술의 발달로 기계 번역의 성능이 사람의 85~90%까지 높아졌다는 평가도 있지만, 맥락을 완전히 이해하거나 감정이 드러나는 글을 번역하는 경우 사람을 따라 잡기에는 아직 역부족이다.

이렇게 번역기가 그 의미를 제대로 이해하지 못하는 이유는 무엇일까?

기계 번역에 오류가 있는 경우를 구체적으로 살펴보면 '게임을 하다, 재밌게'처럼 순서를 고려하지 않고 단어별로 번역하거나 먹는 '배'를 'boat'로 옮기는 경우도 허다했다. 영어를 제대로 해석하려면 문장 전체를 읽고 의미를 파악해야 하는데, 신경망 학습이 적용되기 이전의 번역기는 문장을 구 단위로 쪼개 해석했기 때문에 이러한 오류가 발생했다. 요즘은 기계 번역에 신경망 학습이 적용되어 많은 부분이 개선되었다.

신경망 학습에서 기계 번역은 어떤 식으로 동작할까?



▲ 관련있는 단어는 가까이, 그렇지 않은 용어는 멀리 둔다.

[출처: 수학동아 2017년 2월호]

기계 번역에서 오류가 발생하는 이유를 생각해 보자.

무엇을 알아야 할까?

배경 지식

자연어 처리

인공지능 기술이 발전하면서 사람과 비슷한 수준으로 말하거나 텍스트를 보여 주는 모델이 등장하고 있다. 대표적인 예시로는 사람이 키워드만 보내도 대화형 응답을 제공한다. 이처럼 인공지능이 사람의 말을 이해할 수 있는 능력의 기반은 자연어 처리기술이다.

1. 자연어 처리 적용 형태

컴퓨터는 우리가 일상생활에서 사용하는 자연어의 의미를 분석하여 목적에 맞게 처리한다. 이러한 자연어 처리가 적용된 대표적인 형태는 기계 번역, 대화 시스템(챗봇), 워드 클라우드 등이 있다.

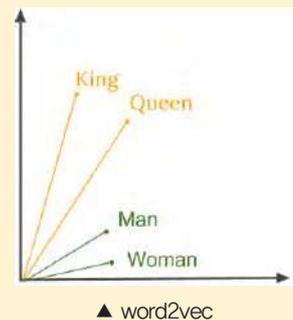
자연어 처리 적용 형태	내용
기계 번역	기계 번역(Machine Translation)은 컴퓨터로 인간이 사용하는 자연 언어를 다른 언어로 번역하는 일을 말하며, 자동 번역이라고도 한다. 대표적인 인공지능 번역기에는 구글 번역, 파파고, 플리토 등이 있다.
챗봇	챗봇(Chatbot)이란 인공지능을 적용한 모바일 채팅 앱, 대화하는 로봇으로 사람이 키워드만 보내도 대화형 응답을 제공한다.
워드 클라우드	워드 클라우드(Word Cloud)는 메타 데이터에서 얻어진 태그들을 분석하여 중요도나 인기도 등을 고려하여 시각적으로 늘어 놓은 것이다.

2. 자연어 처리 과정

- 형태소(Morphological) 분석: 입력된 문자열을 형태소라는 의미를 갖는 최소단위로 쪼갬다.
- 구문(Syntax) 분석: 문장 내에서 주어와 동사 등의 구성을 해석하는 것이다.
- 의미(Semantic) 분석: 구문 분석 단계에서 생긴 의미적 모호성을 해소하고 의미를 파악하는 것이다. 이 과정에서 개체명 인식(named-entity recognition)으로 문장 속에서 사람, 단체, 장소, 시간 등의 개체를 인식하여 분류한다.
- 화용(Pragmatic) 분석: 화자 질문 의도를 추출하고 화자가 원하는 행위를 분석하는 것이다.

3. 임베딩과 word2vec

컴퓨터는 자연어를 직접적으로 처리할 수 없기 때문에 단어를 숫자나 벡터로 변환해야 한다. 임베딩(Embedding)은 단어나 문장을 수치화해서 벡터 공간에 표현하는 것을 말한다. word2vec는 두 개의 유사한 단어가 벡터 공간에 매우 가깝게 배치된 거의 유사한 벡터로 표현된다는 것을 의미한다. 비슷한 단어가 벡터 공간에 밀접하게 배치된다. 오른쪽 그림에서 비슷한 단어인 King과 Queen은 비슷한 공간에 배치된다.



2 기계 번역 체험하기

자연어 처리의 대표적인 형태는 기계 번역이다. 인공지능의 언어 처리를 이해하기 위해 구글 번역, 네이버 파파고, 플리토(Flitto)를 이용하여 같은 문장을 어떻게 번역하는지 비교해 보자.

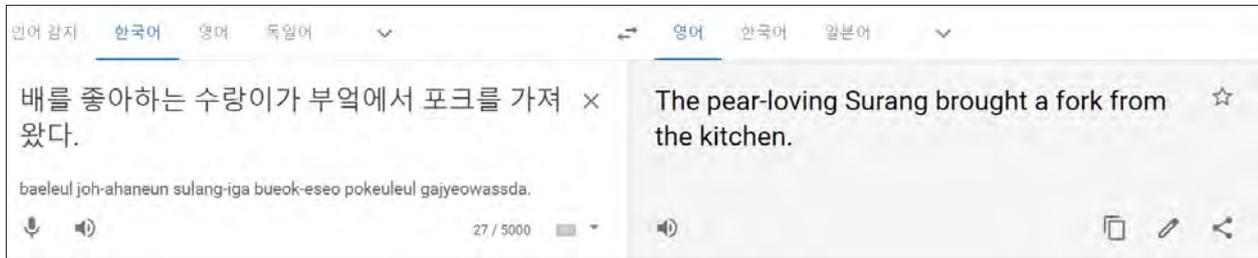
1 한국어-영어 번역하기

다음 한국어 문장을 영어로 번역해 보자. 이 문장에는 ‘배’라는 다의어가 있기 때문에 문장에서 이 단어의 의미 해석이 필요하다.

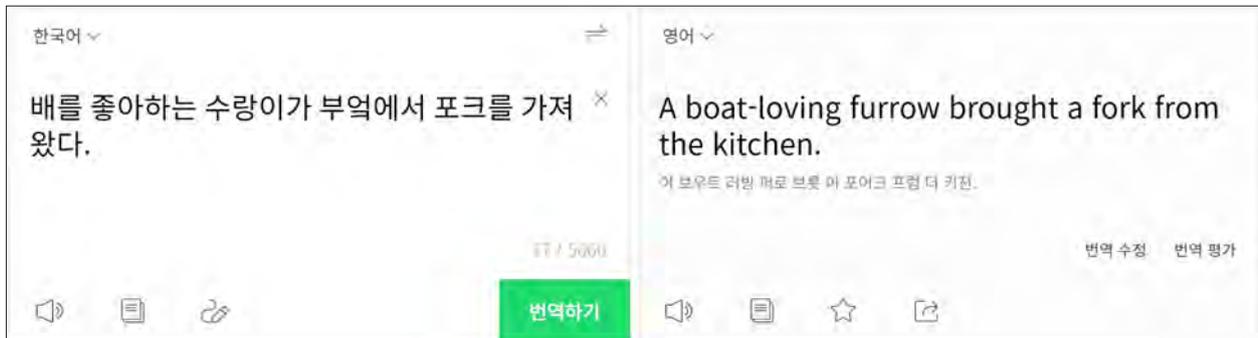
배를 좋아하는 수랑이가 부엌에서 포크를 가져왔다.



1 구글 번역기를 이용하여 한국어 문장을 영어로 번역해 보자.



2 네이버 파파고를 이용하여 한국어 문장을 영어로 번역해 보자.



3 플리토를 이용하여 한국어 문장을 영어로 번역해 보자.



④ 세 가지 번역기의 번역 결과를 비교해 보자.

번역기	번역 결과
구글 번역	
파파고	
플리토	

⑤ 세 가지 번역 결과의 정확도는 어느 정도이고, 어떤 한계가 있는지 작성해 보자.

2 영어-한국어 번역하기

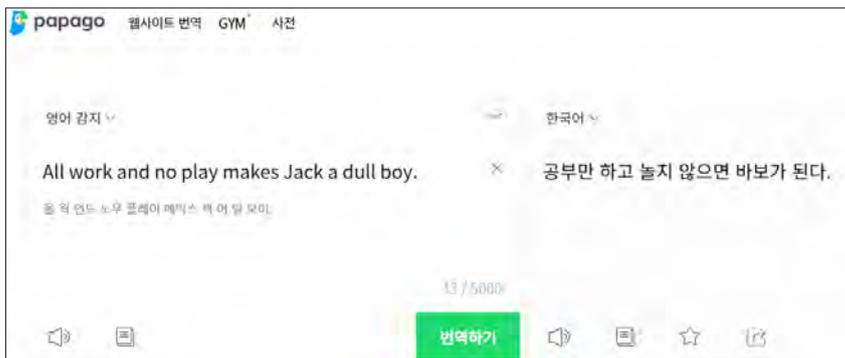
아래 영어 문장을 한국어로 번역해 보자. 이 문장에서 Jack은 특정한 인물을 가리키는 것이 아니며, 성별에 상관 없이 쓰인다.

All work and no play makes Jack a dull boy.

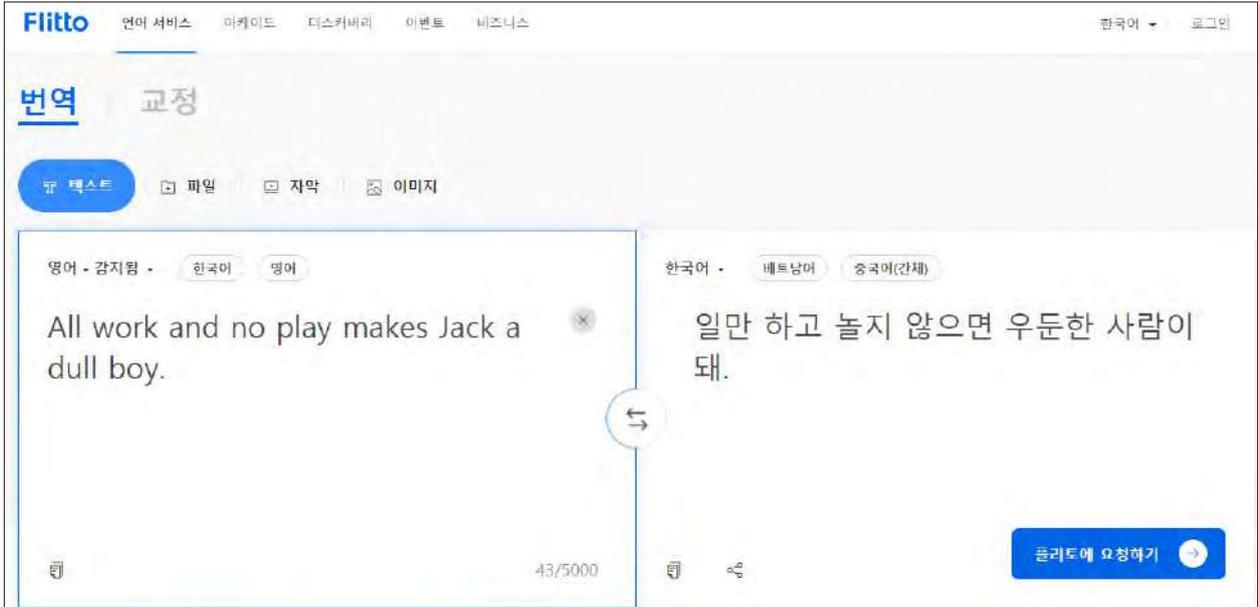
① 구글 번역기를 이용하여 영어 문장을 한국어로 번역해 보자.



② 네이버 파파고를 이용하여 영어 문장을 한국어로 번역해 보자.



③ 플리토를 이용하여 영어 문장을 한국어로 번역해 보자.



④ 세 가지 번역기의 번역 결과를 비교해 보자.

번역기	번역 결과
구글 번역	
파파고	
플리토	

⑤ 위 문장은 어떻게 번역하는 것이 옳을지 의견을 나누어 보자.

3 더 번역해 보기

① 다른 문장을 입력하여 한국어→영어, 영어→한국어로 번역해 보자.

- 문장: _____
- 기대하는 번역 결과: _____
- 문장을 선정한 이유: _____

② 세 가지 번역기의 번역 결과를 비교해 보자.

번역기	번역 결과
구글 번역	
파파고	
플리토	

③ 위 번역이 보다 정확해지려면 우리는 어떤 노력을 해야 할까?

4 분석하기

① 다른 인공지능 번역기의 정확도는 어느 정도인지 평가해 보고, 한계는 무엇인지 써 보자.

	번역기	정확도	세부 평가
1	기계 번역1 ()	☆☆☆☆☆	
2	기계 번역2 ()	☆☆☆☆☆	
3	기계 번역3 ()	☆☆☆☆☆	

② 번역의 정확도를 높이기 위해서는 무엇을 해야 할지 생각해 보자.

2 준비한 말뭉치로 워드 클라우드를 만들어 보자.

- ① 텍스트 입력창에 준비한 말뭉치를 입력하고, **워드클라우드 만들기** 를 클릭하여 결과를 확인한다.

워드클라우드 생성기 3.4

글자색 rainbow 폰트 나눔고딕 T 폰트미러보기

배경색 [White] [Black] [Yellow] [Red] [Grey] [Blue] [Pink] [Purple] [Green]

마스크 [Heart] [Star] [Circle] [Diamond] [Arrow] [X] [Camera] [Checkmark]

크기 직접입력 500px X 500px

단어수 300개

키워드

텍스트

의 바다로
가자) ▲2019년도 '그대만큼 사랑스러운 사람을 본 일이 없다'(김남조의 '편지')
▲2020년
도 '너무 맑고 초롱한 그 중 하나 별이여'(박두진의 '별밭에 누워') 2021년 '맑고
많은 사람

워드클라우드 만들기 저장&공유

- ② 만들어진 워드 클라우드를 이미지로 저장하거나 공유할 수 있다.

워드클라우드 다운로드

공유하기

생성일	2021-01-25 11:13:40
키워드	
글자색	rainbow
폰트	나눔고딕
배경색	white
마스크	circle
크기	500*500

[마스크] 형태를 바꾸어 저장하여 친구들과 비교해 보세요.

③ 다른 워드 클라우드 생성기(<https://www.wordclouds.com/>)로 같은 말뭉치를 사용하여 워드 클라우드를 만들어 보자.

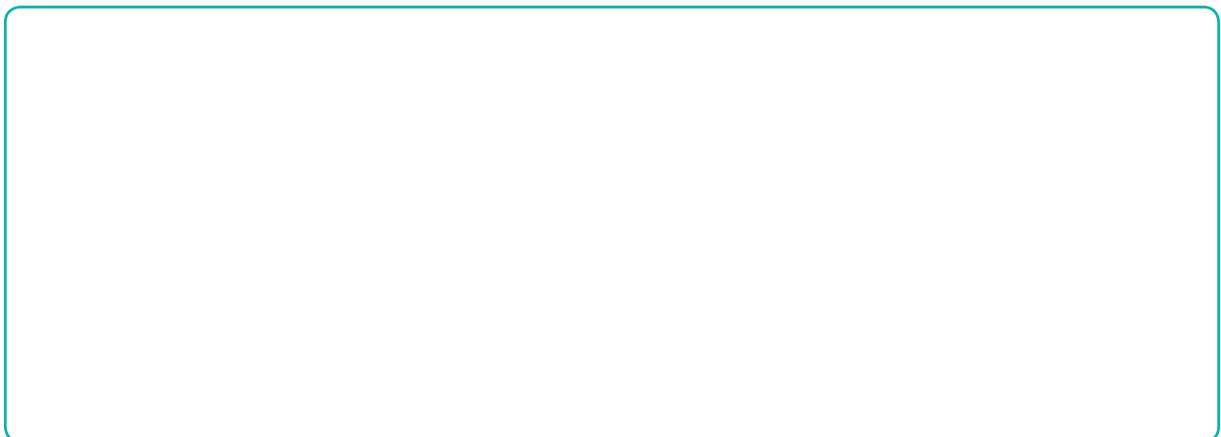
말뭉치를 직접 붙여넣기하거나 파일로 저장하여 워드 클라우드를 만들 수 있다.



사용하고자 하는 말뭉치를 .txt 파일로 저장한 후 확인해 보세요.



④ 워드 클라우드를 만들기 위해 인공지능은 어떤 처리를 하는지 그 과정에 대해 알아보자.



4 정리하기

1 이 활동을 통해 알게 된 점을 작성해 보자.

- 인공지능은 언어를 처리하기 위해 _____ 한다.

- 인공지능이 언어를 이해하는데 한계는 _____ 이다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
준비	활동 전 배경 지식을 충분히 이해하였는가?					
자연어 처리	인공지능의 자연어 처리 과정을 이해하였는가?					
언어 이해의 한계	인공지능이 언어를 이해하는 데 어떤 한계가 있는지 설명할 수 있는가?					

평가는 자신 또는 동료 평가로 진행할 수 있어요.





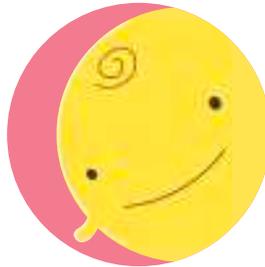
1960년대



ELIZA

미국 MIT에서 엘리자를 개발했다.
엘리자는 문장 구조를 파악한 다음에 준비된 답을 하는 방식이다.

2000년대



심심이

2002년에 개발된 심심이는 우리나라에서 가장 유명한 챗봇이다.
일상적인 대화를 나눌 수 있었다.

2010년 이후



인공지능 개인 비서

스마트폰의 등장과 함께 챗봇도 발달하기 시작했다.
인공지능 개인 비서는 의미 없는 이야기나 농담까지 이해할 수 있게 되었다.



책이나 논문 등 문어체로 쓰인 글을 공부한 챗봇은 자연스럽게 대화하지 못했다.

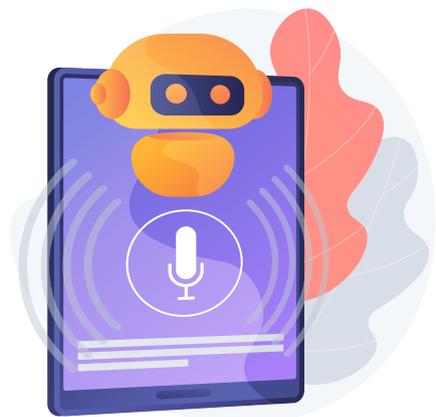


스마트폰이 등장한 이후 사람들의 메시지 자료를 학습한 챗봇은 진짜 사람처럼 말했다.



기본적인 응대 등 온라인 상담의 일부를 챗봇이 하고 있다. 대화의 상대방이 사람인지 챗봇인지 구분하기 어려울 정도이다.

현재



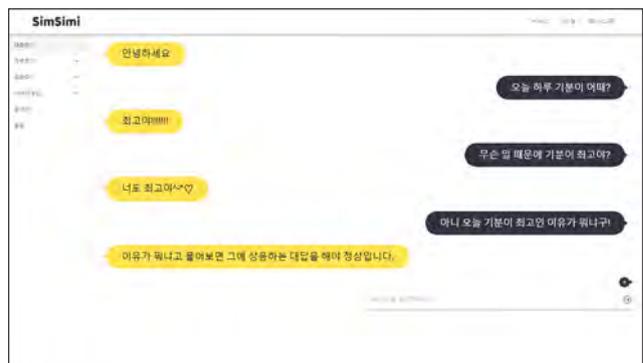
최초의 챗봇은 1966년 미국 MIT 인공지능 연구소의 조셉 바이젠바움(Weizenbaum)이 개발한 엘리자(ELIZA)이다. 엘리자라는 컴퓨터 프로그램에 질문을 입력하면 대답이 화면에 나타났다. 엘리자는 다양한 대화를 할 수 있었는데, 의사 역할을 맡아 환자를 응대하는 것이 가장 잘 알려졌다. 엘리자는 인공지능 기술의 완성도를 평가하는 테스트를 통과했을 만큼 자연스럽게 대화했다. 엘리자의 원리는 문장 구조를 파악해 미리 준비된 대답을 꺼내 놓는 방식이었다.

Talk to Eliza by typing your questions and answers in the input box.



▲ 엘리자와의 대화(로 시작하는 문장이 엘리자)

우리나라에서 가장 유명한 챗봇은 2002년 개발된 '심심이'이다. 마이크로소프트의 MSN 메신저 서비스에 친구로 등록하면 병아리 모양의 로봇 캐릭터와 채팅으로 일상의 소소한 대화를 나눌 수 있는 서비스이다. 심심이는 사람들의 말을 분석한 후 미리 준비된 답을 하는 기초적인 형태였다. 그럼에도 심심이는 당시 큰 인기를 누렸고, 지금도 채팅뿐 아니라 음성 대화로도 영역을 확장하고 있다.



▲ 심심이와의 대화(왼쪽이 심심이)

스마트폰 음성 비서 기능이 등장하면서 챗봇은 재미를 넘어 실제 기술로 주목받게 된다. 2011년 애플은 스마트폰인 아이폰을 발표하면서 '시리(Siri)'라는 인공지능 비서 기능을 선보였다. 시리는 말로 기기를 제어하고, 원하는 정보를 검색할 수 있게 해주었다. 이전까지 음성 인식 기술은 정해진 명령어만 알아듣는 것이 일반적이었다. 이제 챗봇이 우리가 쓰는 일상의 언어, 즉 자연어를 알아들을 수 있게 되었고, 우리가 하는 의미 없는 이야기나 농담까지 이해할 수 있다.





인식 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

1 배운 내용을 비주얼 씽킹으로 표현하고, 친구들과 비교해 보자.

주제				
항목				
관련 내용		→	→	→
비주얼 씽킹으로 표현하기				
배우고 느낀 점				

- [작성 방법]**
1. 항목에는 인식의 4가지 활동을 작성한다.
 2. 관련 항목에는 주요 학습 요소를 작성한다.
 3. 알게 된 사실을 글과 그림으로 표현한다.

주제	인공지능은 어떻게 주변 환경을 인식할까?			
항목	센서와 인식	이미지 인식	음성 인식	언어 이해
관련 내용		컴퓨터 비전 이미지 캡처 이미지 분류		
비주얼 씽킹으로 표현하기				
배우고 느낀 점				

학습 과정을 되돌아보며 자신이 알게 된 내용을 글과 그림으로 표현합니다.



2 프로젝트를 진행하면서 인간과 인공지능이 인식의 관점에서 어떻게 다른지 비교하여 작성해 보자.

인간

인공지능

3 앞으로 인간과 인공지능이 공존하는 미래의 모습을 친구와 함께 이어 쓰기로 작성해 보자.

주제: 인공지능과 함께 하는 미래의 생활에서 우리 인간은 어떤 노력을 해야 할까?

(모듬원1)

(모듬원2)

(모듬원3)

(모듬원4)

(모듬원1)

(모듬원2)

이어쓰기는 각자의 생각을 작성하는 것이 아니라, 하나의 주제에 대해 앞의 글에 이어 글을 작성하는 것이예요.



Ⅲ 데이터

1 단계 핵심 개념 알기

- 데이터의 역할
- 데이터 시각화
- 데이터 가공



2 단계 창의융합적 사고로 해결하기

- 활동 1 사람의 기대 수명은 얼마나 될까?
- 활동 2 다른 사람들은 인터넷으로 무엇을 할까?
- 활동 3 올 겨울 유행할 외투는 무엇일까?
- 활동 4 유명 롤러코스터를 살펴볼까?
- 활동 5 남녀공학 여부에 따른 나의 성적 평가는?

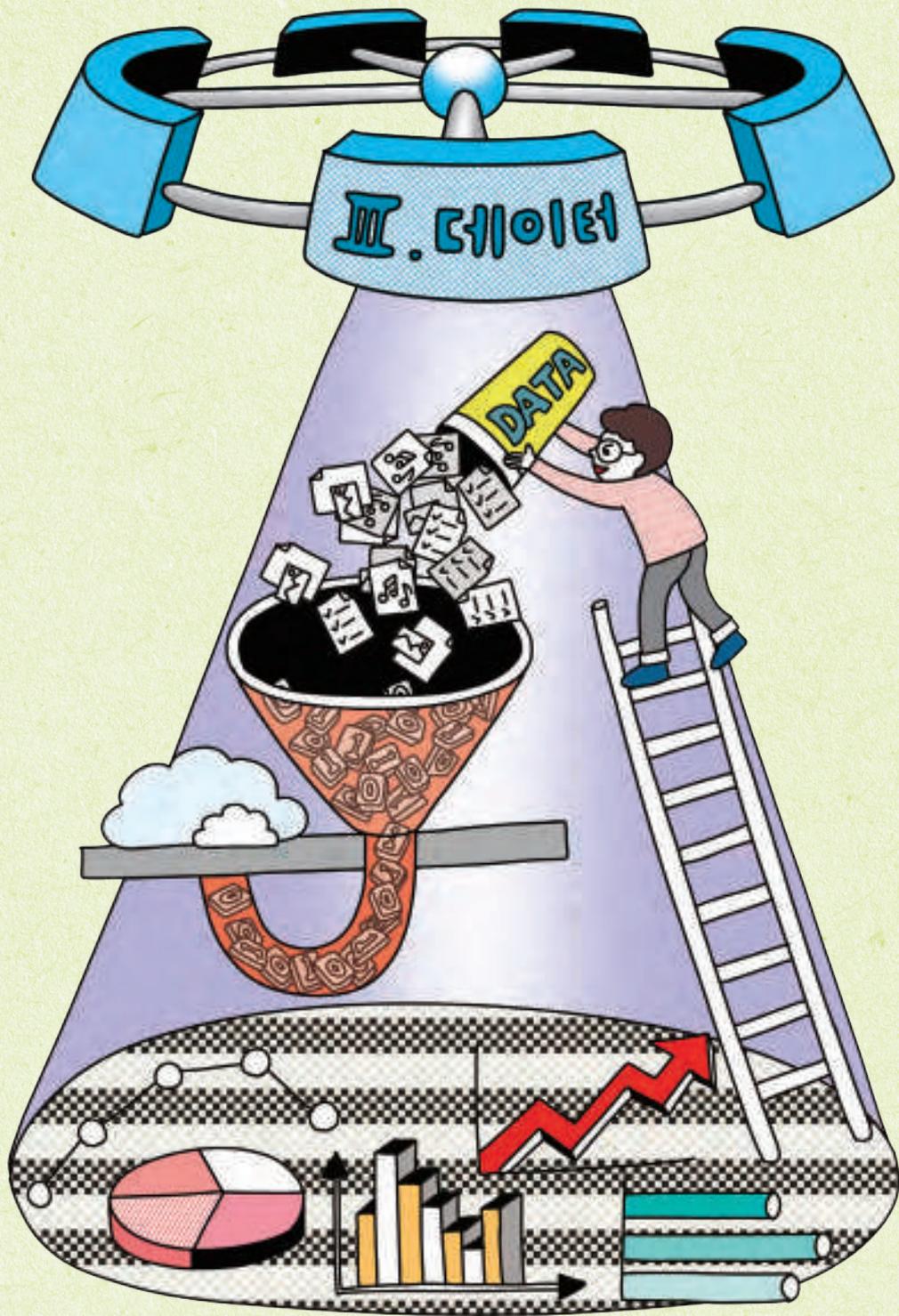


3 단계 핵심 개념 알기

- 마인드맵으로 정리하기

④ 데이터 단원에서는 무엇을 배울까요?

인공지능 활용에 있어 데이터의 역할과 특성을 이해하고 데이터를 문제해결에 맞게 처리하여 시각화합니다.





1단계 핵심 개념 알기

데이터 이해와 활용

✔ 데이터로 어떤 문제를 해결할 수 있을까?

데이터는 인공지능이 점점 똑똑해지도록 만드는 재료가 된다. 그러므로 데이터가 인공지능에서 어떤 의미를 가지는지, 데이터는 어떤 형태이고 어떻게 처리해야 하는지 등을 이해할 수 있어야 한다. 본 단원에서는 데이터에 대해 이해하고, 데이터를 처리하는 방법, 데이터를 시각화하는 방법에 대해 알아본다.

1 데이터의 역할

Q 개념 탐색 데이터로 무엇을 할 수 있을까?

데이터 역할을 이해하고 속성의 의미를 알면 이 데이터를 이용하여 새로운 정보들을 찾고 미래를 예측하는 등 다양한 문제를 해결할 수 있다.

- 사람은 다양한 경험을 통해 배우고 진화하지만 인공지능은 데이터를 통해 학습하고 발전한다.
- 데이터에는 어떤 데이터인지 특징을 나타내는 다양한 속성이 있으며, 이 속성은 문제 해결에 사용된다.

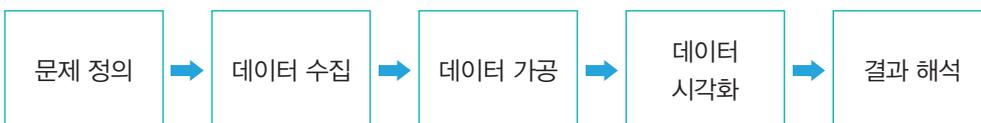
예) 성별에 따른 수명 데이터에는 나이, 여성, 남성이라는 속성이 존재한다.

2 데이터 시각화

Q 개념 탐색 데이터를 시각화하면 편리한 점은 무엇일까?

데이터를 시각화함으로써 데이터가 의미하는 바를 한눈에 파악할 수 있고 새로운 정보를 찾을 수도 있다.

- 예) • 꽃이 피는 식물과 꽃이 피지 않는 식물의 성장에는 어떤 차이가 있는지 알아보기 위해 식물 성장 데이터를 시각화한다.
• 올 겨울 유행할 외투는 무엇인지 파악하기 위한 검색 데이터를 시각화한다.



▲ 데이터 분석 과정

“데이터의 속성을 추출하여 인공지능으로 문제를 해결할 수 있는 형태로 가공하고 처리한 후, 데이터가 알려 주는 정보를 파악한다.”

3 데이터 가공

Q 개념 탐색 데이터를 처리하려면 어떻게 가공해야 할까?

인공지능이 데이터로 문제를 해결할 수 있도록 하려면 수집한 데이터의 유형과 값을 살펴보고 처리할 수 있게 가공해야 한다.

1 인공지능이 사용하는 데이터의 유형

- 정형 데이터: 의미를 파악하기 쉽도록 규칙에 따라 정리된 데이터이다.
- 비정형 데이터: 의미를 파악하기 쉽지 않고 규칙을 갖지 않은 데이터이다.

2 데이터들을 수집하는 방법

- 공공 데이터: 공공기관에서 수집·배포하는 데이터로 무료로 사용할 수 있다.
- 민간 데이터: 수익을 목적으로 유료로 제공하는 데이터로 다양한 데이터에 접근할 수 있다.
- 직접 수집(설문 등): 필요와 목적에 맞는 데이터를 직접 수집할 수 있지만 비용과 시간이 많이 든다.

3 데이터 가공

- 필요한 데이터 추출: 해결하려는 문제에 맞는 속성을 데이터 셋*에서 추출한다. (*데이터 셋: 데이터들의 집합)
- 결측값 처리: 결과에 영향을 미치지 않도록 값이 비어 있는 데이터를 처리한다.

인공지능으로 문제를 해결하려면 수집한 데이터를 데이터 유형에 맞게 가공해서 사용해야 합니다.



활동 미리 보기

핵심 개념과 관련된 활동은 어떤 내용으로 전개될까요?

활동 1

사람의 기대 수명은 얼마나 될까?

- 데이터의 의미
- 데이터의 속성과 역할

활동 2

다른 사람들은 인터넷으로 무엇을 할까?

- 데이터 시각화

활동 3

올 겨울 유행할 외투는 무엇일까?

- 데이터 분석

활동 4

유명 롤러코스터를 살펴볼까?

- 정형 데이터
- 비정형 데이터

활동 5

남녀공학 여부에 따른 나의 성적 평가는?

- 데이터 수집 및 정제
- 데이터 분석



2단계 창의융합적 사고로 해결하기

활동 1 사람의 기대 수명은 얼마나 될까?

- ▶ 사람의 기대 수명을 알아보는 데이터로 시각화 작업을 하고 데이터의 의미와 특징 등 속성을 설명할 수 있다.

1 문제 이해하기

다음 글을 읽고, 사람의 기대 수명을 예측해 보자.

분열의 시기에 중국을 하나로 만든 사람이 진나라 36대 왕, 진시황이다. 13세에 왕이 된 진시황은 이사를 재상으로 등용해 강력한 부국강병책을 추진하였다. 진시황은 7개의 큰 나라들이 각축을 벌이던 혼란을 종식시키고, 기원전 221년 중국을 통일하였다.

평화의 시대가 오자 진시황은 늙지 않고 죽지 않는 약을 찾고자 하였다. 신비의 약을 바라던 진시황은 효험 있다는 약을 먹어도 신통치 않자 속이는 자는 사형에 처한다는 법까지 만들었다. 약을 좋아한 진시황의 약물 집착은 결국 수은 중독으로 막을 내렸다.



이처럼 사람은 수명에 관해 많은 관심과 기대를 가지고 살아간다. 어떤 사람은 100세 이상 살기도 하고 또 어떤 사람들은 오래 살지 못하고 생을 마감하기도 한다.

그렇다면 일반적으로 특정 나이의 사람이 얼마나 더 살 수 있을지 예상되는 수명을 알아보려면 어떻게 해야 할까? 과연 사람은 얼마나 살 수 있을지 알아보기 위해 기대 수명 데이터로 시각화해 보고 데이터, 데이터의 속성은 무엇인지 살펴보고 데이터의 역할을 파악해 보자.

이번 활동에서 해결해야 할 문제는 무엇인지 써 보자.

.....

.....

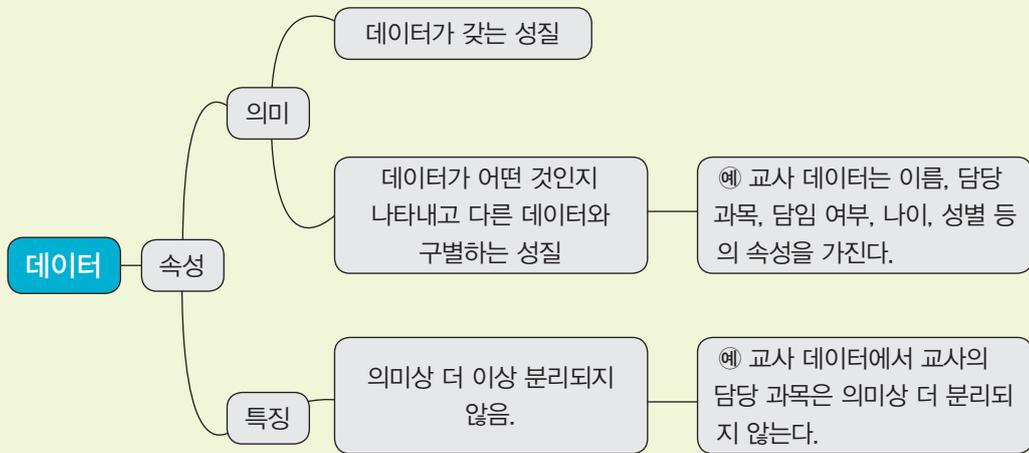
무엇을 알아야 할까?

배경 지식

데이터의 속성

인공지능은 매우 방대한 양의 데이터를 이용하기 때문에 데이터의 중요성이 강조된다. 하지만, 정확하지 않거나 의미 없는 데이터는 인공지능의 성능을 떨어뜨리기도 한다. 인공지능한테 필요한 데이터인지는 속성을 이해하는 것에서부터 시작한다. 사람이 좋고 의미 있는 경험을 통해 올바르게 성장하는 것처럼 인공지능에게는 양질의 데이터를 공급하는 것이 미래를 예측하고 다양한 문제를 해결하는 데 중요한 역할을 한다. 사람은 다양한 경험을 통해 배우고 똑똑해지는 반면, 인공지능은 데이터를 통해 똑똑해지며 이 데이터들을 토대로 새로운 정보를 찾고 미래를 예측하는 등 다양한 문제를 해결할 수 있다.

데이터에는 다양한 속성이 존재하며 이 속성은 문제 해결에 사용된다. 예를 들어, 성별에 따른 수명 데이터에는 나이, 여성, 남성이라는 속성이 존재하며 몇 살까지 살았는지에 대한 데이터를 바탕으로 여성과 남성의 수명을 비교하거나 앞으로 얼마나 살 수 있는지를 예측할 수 있다.



Period Life Table, 2013				
Period life expectancy at a given age (120 cases)				
ind... ...dex	age	male	female	both sexes
1	0	76.28	81.05	78.67
2	1	75.78	80.49	78.14
3	2	74.82	79.52	77.17
4	3	73.84	78.54	76.19
5	4	72.85	77.55	75.2

▲ 기대 수명 데이터

2 분석 플랫폼 이해하기

데이터 분석 도구인 코답(CODAP) 플랫폼 사용 방법을 알아보자.

1 코답(CODAP) 플랫폼 접속하기

주소 창에 URL을 입력하여 접속한다. 이때 크롬() 브라우저에서 접속해야 오류 없이 사용할 수 있다.

<https://codap.concord.org/>

2 데이터가 준비되어 있는 경우 사용 방법

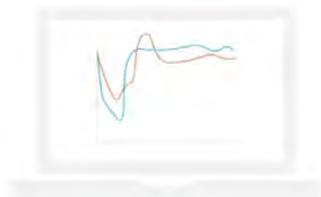
사전에 데이터 파일을 가지고 있다면 분석만 실시하면 되므로, 사이트의 우측 상단의  버튼을 이용하여 프로그램을 실행한다.

코답(CODAP)은 영문으로 된 플랫폼이지만, 크롬에서 마우스 오른쪽 버튼 클릭하여 '한국어로 번역'으로 변환하면 사용하기 편리합니다.



3 코답에서 제공하는 예시 데이터를 사용하는 방법

프로그램 실행 전 예시 데이터를 선택하고 분석하는 것이 편리하다. 따라서 사이트 중간 위치의 'Sample CODAP Datasets'을 이용하는 것이 데이터를 선택하는 데 편리하다.



Try CODAP

CODAP is easy to use and runs in your web browser. CODAP is (and always will be) free. Share your data with others and bring it to life!



Get the Code

Learn how to create your own plugin in CODAP. Explore our open-source GitHub repository and start incorporating it into your work today.



Sample CODAP Datasets

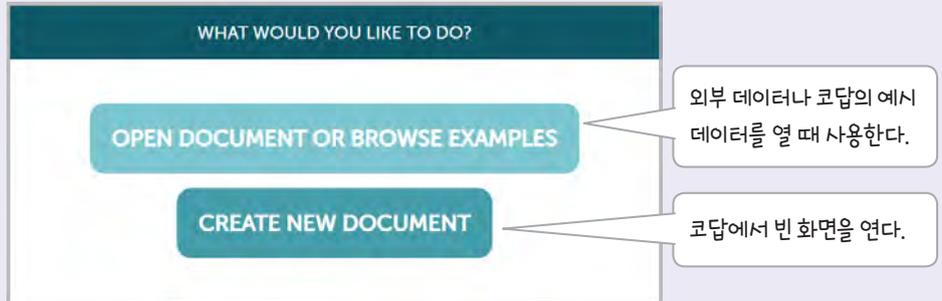
Browse free CODAP example documents. We've curated datasets for first-time users. Explore to find classroom activities and other downloadable resources.

도움 자료

코답에서 제공하는 예시 데이터를 불러오는 방법

데이터가 준비되지 않은 상태에서 코답의 예시 데이터를 사용할 수 있다. 우측 상단의 **Try CODAP** 버튼을 클릭하면, 텍스트로 구성된 데이터 목록만 보고 데이터를 선택해야 하므로 다음 과정을 거쳐 데이터를 불러온다.

❶ 데이터를 어떻게 불러올 것인지 선택한다.



❷ 목록에서 원하는 데이터를 선택하거나 파일을 직접 선택하여 불러온다.



※ 코답(CODAP)

- 코답(CODAP)은 데이터 분석을 프로그램 설치 없이 온라인에서 바로 할 수 있도록 만든 웹 기반 플랫폼이다.
- 웹 브라우저에서 바로 실행할 수 있으며 다양한 예시 데이터들까지도 무료로 제공되고 있어 데이터 분석 활동이 쉽다.



3 문제 해결하기

문제 해결에 필요한 데이터 셋에서 필요한 속성을 찾아보자.

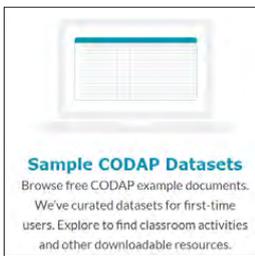
1. 데이터 속성 파악하기

기대 수명 데이터 셋으로 데이터의 속성을 살펴보자.

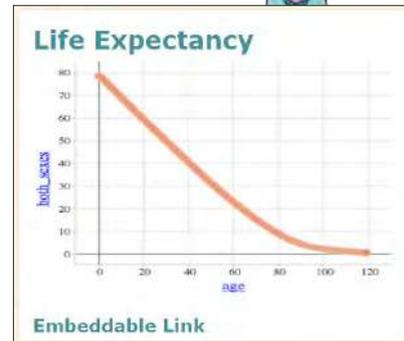
1 데이터 불러오기

플랫폼에서 제공하는 데이터를 불러오기 위해서 'Sample CODAP Datasets'을 클릭하고, 'Life Expectancy'를 선택한다.

분석에 사용할 데이터는 온라인 또는 데이터 제공 사이트, 설문 조사 등을 통해 수집합니다. 본 단원의 목표가 데이터를 이해하는 것이 핵심이므로 데이터 시각화 플랫폼에서 제공하는 데이터를 사용합니다.



▲ 코답 예시 데이터 셋 리스트



▲ 기대 수명 데이터 셋

2 데이터 수집하여 테이블 만들기

아래 테이블에서 데이터의 속성을 찾고 속성의 의미를 알아보자.

Untitled Document UNSAVED

Tables Graph Map Slider Calc Text Plugins

Period life expectancy at a given age (120 cases)				
index	age	male	female	both sexes
1	0	76.28	81.05	78.67
2	1	75.78	80.49	78.14
3	2	74.82	79.52	77.17
4	3	73.84	78.54	76.19
5	4	72.85	77.55	75.2
6	5	71.87	76.56	74.22
7	6	70.88	75.57	73.22
8	7	69.89	74.58	72.24
9	8	68.90	73.59	71.25
10	9	67.91	72.60	70.26
11	10	66.92	71.61	69.27
12	11	65.93	70.62	68.28
13	12	64.94	69.63	67.29
14	13	63.95	68.64	66.30
15	14	62.96	67.65	65.31
16	15	61.97	66.66	64.32
17	16	60.98	65.67	63.33
18	17	59.99	64.68	62.34
19	18	59.00	63.69	61.35
20	19	58.01	62.70	60.36
21	20	57.02	61.71	59.37
22	21	56.03	60.72	58.38
23	22	55.04	59.73	57.39
24	23	54.05	58.74	56.40
25	24	53.06	57.75	55.41
26	25	52.07	56.76	54.42
27	26	51.08	55.77	53.43
28	27	50.09	54.78	52.44
29	28	49.10	53.79	51.45
30	29	48.11	52.80	50.46
31	30	47.12	51.81	49.47
32	31	46.13	50.82	48.48
33	32	45.14	49.83	47.49
34	33	44.15	48.84	46.50
35	34	43.16	47.85	45.51
36	35	42.17	46.86	44.52
37	36	41.18	45.87	43.53
38	37	40.19	44.88	42.54
39	38	39.20	43.89	41.55
40	39	38.21	42.90	40.56
41	40	37.22	41.91	39.57
42	41	36.23	40.92	38.58
43	42	35.24	39.93	37.59
44	43	34.25	38.94	36.60
45	44	33.26	37.95	35.61
46	45	32.27	36.96	34.62
47	46	31.28	35.97	33.63
48	47	30.29	34.98	32.64
49	48	29.30	33.99	31.65
50	49	28.31	33.00	30.66
51	50	27.32	32.01	29.67
52	51	26.33	31.02	28.68
53	52	25.34	30.03	27.69
54	53	24.35	29.04	26.70
55	54	23.36	28.05	25.71
56	55	22.37	27.06	24.72
57	56	21.38	26.07	23.73
58	57	20.39	25.08	22.74
59	58	19.40	24.09	21.75
60	59	18.41	23.10	20.76
61	60	17.42	22.11	19.77
62	61	16.43	21.12	18.78
63	62	15.44	20.13	17.79
64	63	14.45	19.14	16.80
65	64	13.46	18.15	15.81
66	65	12.47	17.16	14.82
67	66	11.48	16.17	13.83
68	67	10.49	15.18	12.84
69	68	9.50	14.19	11.85
70	69	8.51	13.20	10.86
71	70	7.52	12.21	9.87
72	71	6.53	11.22	8.88
73	72	5.54	10.23	7.89
74	73	4.55	9.24	6.90
75	74	3.56	8.25	5.91
76	75	2.57	7.26	4.92
77	76	1.58	6.27	3.93
78	77	0.59	5.28	2.94
79	78	0.60	4.29	1.95
80	79	0.61	3.30	0.96
81	80	0.62	2.31	0.97
82	81	0.63	1.32	0.98
83	82	0.64	0.33	0.99
84	83	0.65	0.34	0.99
85	84	0.66	0.35	0.99
86	85	0.67	0.36	0.99
87	86	0.68	0.37	0.99
88	87	0.69	0.38	0.99
89	88	0.70	0.39	0.99
90	89	0.71	0.40	0.99
91	90	0.72	0.41	0.99
92	91	0.73	0.42	0.99
93	92	0.74	0.43	0.99
94	93	0.75	0.44	0.99
95	94	0.76	0.45	0.99
96	95	0.77	0.46	0.99
97	96	0.78	0.47	0.99
98	97	0.79	0.48	0.99
99	98	0.80	0.49	0.99
100	99	0.81	0.50	0.99
101	100	0.82	0.51	0.99
102	101	0.83	0.52	0.99
103	102	0.84	0.53	0.99
104	103	0.85	0.54	0.99
105	104	0.86	0.55	0.99
106	105	0.87	0.56	0.99
107	106	0.88	0.57	0.99
108	107	0.89	0.58	0.99
109	108	0.90	0.59	0.99
110	109	0.91	0.60	0.99
111	110	0.92	0.61	0.99
112	111	0.93	0.62	0.99
113	112	0.94	0.63	0.99
114	113	0.95	0.64	0.99
115	114	0.96	0.65	0.99
116	115	0.97	0.66	0.99
117	116	0.98	0.67	0.99
118	117	0.99	0.68	0.99
119	118	1.00	0.69	0.99
120	119	1.00	0.70	0.99

- 수집한 데이터를 불러오면 왼쪽 그림과 같이 테이블이 생성되어 어떤 속성이 포함되어 있는지 확인할 수 있다.
- 속성은 데이터가 어떤 것인지 나타내는 것으로 다른 데이터와 구별하는 성질을 가지고 있다. 데이터 테이블에서는 열의 제목이 속성에 해당한다.

예 age: 현재 나이
 male: 남성의 기대 수명
 female: 여성의 기대 수명
 both sexes: 남녀 기대 수명 평균

3 필요한 데이터 추출하기

기대 수명에 대한 데이터 셋을 이용하여 어떤 문제를 해결할 것인지에 따라 서로 다른 속성을 추출하여 사용한다. 다음 문제를 해결하기 위해 필요한 속성을 적어 보자.

문제 상황	필요한 속성
나이에 따라 사람은 얼마나 더 살 것이라고 기대할까?	
나이에 따라 여성은 얼마나 살 것이라고 기대할까?	
나이에 따라 남성은 얼마나 살 것이라고 기대할까?	
남성의 기대 수명에 따른 여성의 기대 수명은 어떨까?	

나이에 따라 남녀를 구분하지 않고 사람이 얼마나 살지를 알아보려면, 나이(age) 속성과 남녀 합한 기대 수명(both sexes) 속성이 필요합니다.



도움 자료

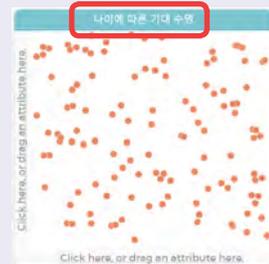
시각화

코담에서 그래프를 그리기 위해서는 다음의 과정을 거친다. 예를 들어, 나이에 따라 사람의 수명에 대한 기대치가 어느 정도 되는지를 알아보기 위한 그래프는 다음과 같은 과정을 통해 작성한다.

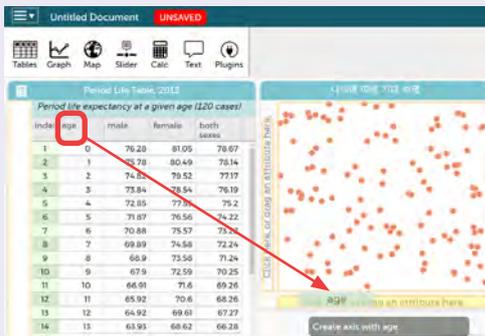
① 상단의 메뉴 중에서 그래프(Graph)를 선택한다.



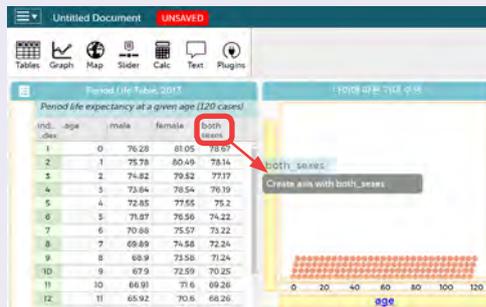
② 임의의 산점도가 생기면 차트 이름을 '나이에 따른 기대 수명'으로 변경한다.



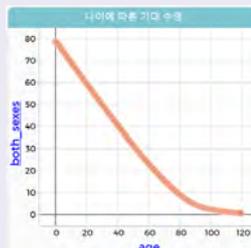
③ 테이블의 나이(age) 속성을 그래프의 x축으로 드래그 앤 드롭(drag&drop)하여 x축을 생성한다.



④ 테이블의 남녀 합한 기대 수명(both sexes) 속성을 그래프의 y축으로 드래그 앤 드롭(drag&drop)하여 y축을 생성한다.



⑤ 사람들이 해당 나이에서 얼마나 더 살 수 있을지 기대하는 그래프가 완성된다.



2. 우리 학급 친구 데이터 셋 만들기 심화 활동(선택)

새로운 테이블을 생성하여 우리 학급 친구들에 대한 데이터 셋을 만들어 보자.

1 주제 정하기

우리 학급 친구들에 대한 데이터 수집 및 분석을 통해 알고 싶은 주제를 정해 보자.

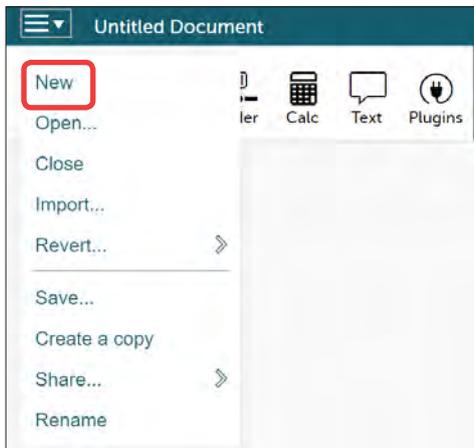
예) 하루 과목별 공부 시간, 좋아하는 음식 등

2 데이터 속성 정하기

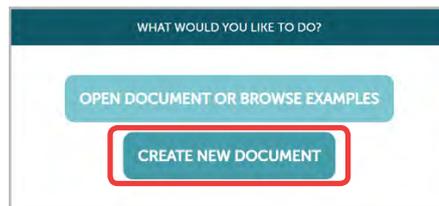
앞에서 정한 주제를 파악하기 위해 데이터를 수집하여 테이블을 만들 때, 속성으로 필요한 것을 적어 보자.

3 데이터 테이블 만들기

① 메뉴()에서 'New'를 선택한다.



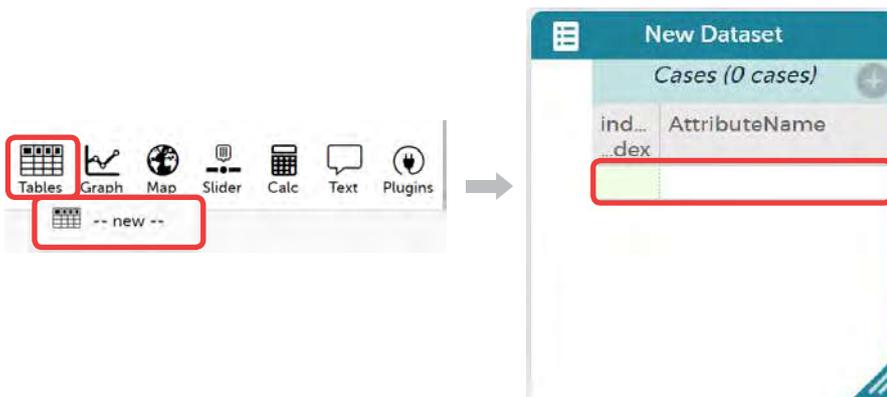
② 'CREATE NEW DOCUMENT'를 클릭한다.



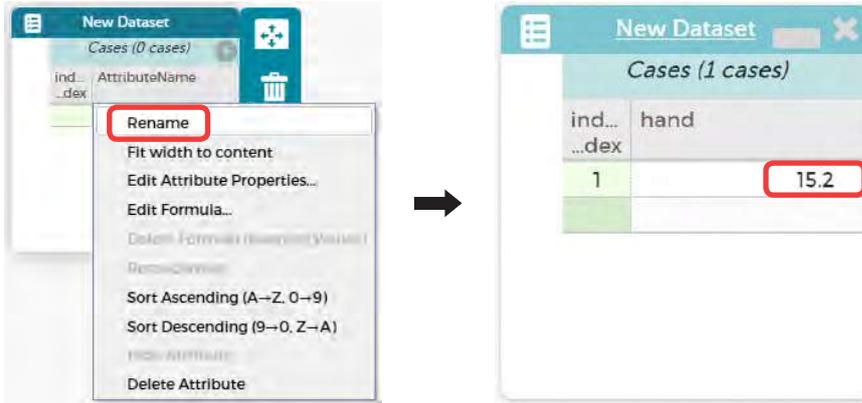
코담에서 테이블을 만드는 방법을 사용해 보세요.



③ Tables를 선택하고, '--new--'를 선택하면 빈 테이블을 생성된다.



④ **+**를 눌러 필요한 개수만큼 속성을 만들고, 속성명을 정한 후 데이터를 입력한다.



앞의 [도움 자료]의 시각화를 이용하여 그래프를 그려 보세요.



4 데이터 시각화하기

입력한 데이터를 이용하여 그래프로 나타내 보자.

4 정리하기

1 활동을 통해, 알게 된 내용을 정리해 보자.

- 데이터란 _____
_____이다.
- 데이터의 속성은 _____
_____이다.
- 데이터의 역할은 _____
_____이다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
데이터 속성	데이터의 속성이 무엇인지 이해하였는가?					
데이터 역할	데이터의 역할이 무엇인지 이해하였는가?					
참여도	활동에 적극적으로 참여하였는가?					

활동 2

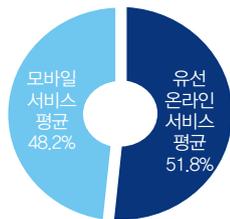
다른 사람들은 인터넷으로 무엇을 할까?

- ▶ 데이터 시각화 형태를 이해하고 사람들의 인터넷상 SNS 관심 분야를 탐색할 수 있다.

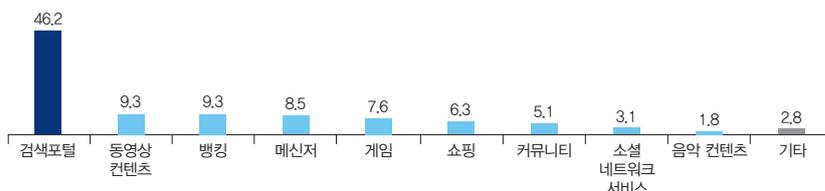
1 문제 이해하기

다음 글을 읽고, 사람들이 인터넷상에서 검색할 때 관심을 갖는 분야가 무엇인지를 알아보자.

2018년 한국인터넷기업협회 설문 조사에 따르면 우리나라 개인당 1일 평균 인터넷 사용 시간은 약 4시간 27분이라고 한다. 인터넷을 사용할 때 유선 온라인 서비스를 사용하는 경우는 약 51.8%이고 모바일 서비스를 사용하는 경우는 약 48.2%이다. 이 시간 중 유선 인터넷 사용자의 약 46.2%가 검색 포털로 정보를 검색하고, 모바일을 통한 인터넷 사용자의 22.7%가 검색 포털로 정보를 검색한다.



[유선·모바일 인터넷 사용 비중]



[주로 이용하는 유선 온라인 서비스]



[주로 이용하는 모바일 서비스]

그렇다면 이처럼 오랜 시간 동안 무엇을 검색하는 것일까? 검색 정보를 토대로 나이나 성별에 따른 사람들의 관심사를 시각화해 보고 그 결과를 분석해 보자.

[출처 | <https://post.naver.com/viewer/postView.nhn?volumeNo=15872365&memberNo=967>]

유선이나 모바일 인터넷 사용에서 공통으로 많이 하는 작업이 검색이라는 것을 알 수 있어요.



이번 활동에서 해결해야 할 문제는 무엇인지 써 보자.

무엇을 알아야 할까?

배경 지식

데이터 시각화

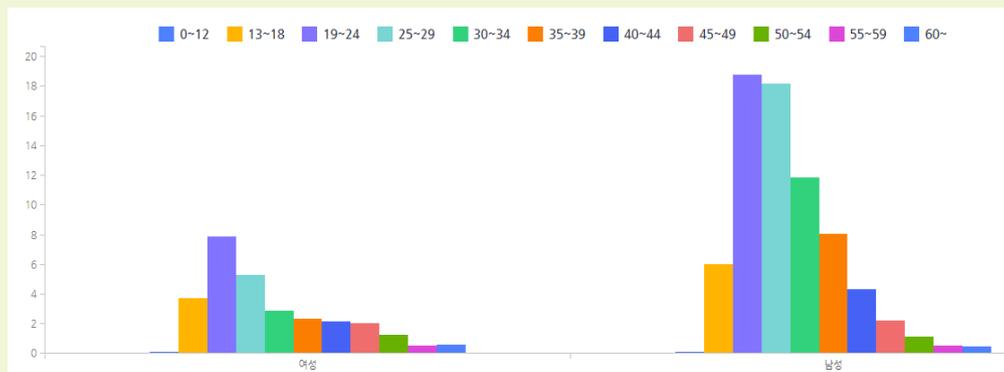
방대한 양의 데이터는 낱개의 데이터와 달리 데이터의 관계와 패턴 분석을 통해 새로운 정보를 만드는 데 사용된다. 하지만, 많은 양의 데이터의 관계나 패턴을 파악하기 어렵기 때문에 시각화를 하면 데이터의 의미를 한 눈으로 파악할 수 있고, 평균 등의 요약보다 전반적으로 데이터의 경향성을 파악할 수 있다.

이번 활동에서는 온라인 데이터 분석 도구를 사용하여 데이터를 분석하고 시각화하여 새로운 정보를 파악한다. 이때 사용하는 온라인 데이터 분석 도구는 프로그램 설치 없이 바로 데이터 분석에 활용할 수 있다. 우리가 사용할 프로그램인 엔트리는, 데이터를 불러와 간단한 조작을 통해 쉽게 그래프로 시각화할 수 있다는 장점이 있다.

No.	연령대	성별	건강-의학	게임	공연-전시	교육-학문	국내여행	드라마
1	0-12	여성	0.03	0.1	0.02	0.07	0.02	0.03
2	0-12	남성	0.01	0.11	0.01	0.07	0.01	0.02
3	13-18	여성	1.74	3.68	1.42	3.87	1.04	2.06
4	13-18	남성	0.7	5.98	0.36	1.92	0.64	1.82
5	19-24	여성	9.4	7.85	7.38	13.66	6.7	8.27
6	19-24	남성	3.6	18.77	2.17	7.25	3.16	5.26
7	25-29	여성	9.6	5.28	9.12	8.77	9.56	9.46
8	25-29	남성	4.53	18.15	3.25	5.87	4.75	4.92
9	30-34	여성	8.88	2.83	10.2	6.78	9.92	10.05
10	30-34	남성	4.23	11.85	3.53	4.21	5.28	4.79
11	35-39	여성	10.35	2.32	14.43	7.71	12.55	11.74
12	35-39	남성	4.58	8.01	4.18	4.02	6.56	5.3

▲ 성별, 연령대별 검색 분야 데이터 예

↓ 그래프로 시각화



▲ 성별, 연령별 IT 분야 검색 그래프 예

2 분석 플랫폼 이해하기

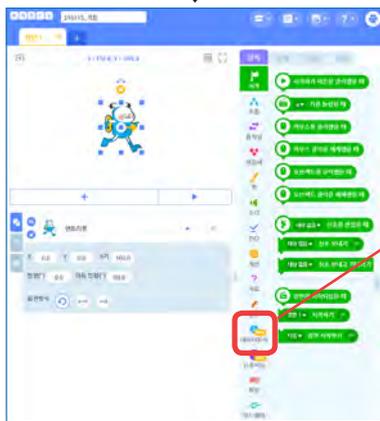
데이터 분석 도구인 엔트리(ENTRY) 플랫폼 사용 방법을 알아보자.

1 데이터 분석을 위한 엔트리(ENTRY) 플랫폼 접속하기

① 주소창에 URL을 입력하여 접속한다.

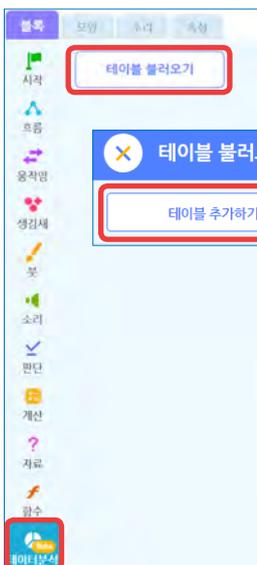
https://playentry.org

② [만들기] - [작품 만들기] - [데이터 분석]으로 이동한다.



③ 데이터 셋을 추가한다.

데이터 분석을 위해 [데이터 불러오기] - [데이터 추가하기] 를 눌러 계절별 기온 데이터를 선택한 다음, [추가] 버튼을 누른다.



2 데이터 셋 다루기

- ① 데이터 셋을 불러오면 테이블 메뉴에서 실제 각 데이터의 속성과 값을 확인할 수 있다.

요약 테이블은 데이터의 특성을 빠르게 파악할 수 있도록 돕는다는 점에서 반드시 살펴보는 것이 좋습니다.

		A	B	C	D	E	F
속성	1	연도	연평균	봄	여름	가을	겨울
값	2	1973	12.4	11.6	24.5	12.9	-1.4
	3	1974	11.4	10.8	22.4	13.0	-0.1
	4	1975	12.6	11.2	23.9	15.5	0.3
	5	1976	11.7	10.9	22.6	12.5	-1.7
	6	1977	12.3	11.8	23.5	15.0	0.7

- ② 데이터 셋에 대한 정보 테이블을 확인할 수 있다. 정보 테이블에는 데이터 속성의 종류와 개수, 평균, 표준편차, 최댓값, 중간값, 최솟값과 같은 기본적인 통계치가 포함되어 있다.

	평균	표준 편차	최댓값	중간값	최솟값
연도	1995.5	13.28	2018	1995	1973
연평균	12.52	0.57	13.6	12.5	11.2
봄	11.79	0.72	13.4	11.7	10.6
여름	23.73	0.82	25.4	23.6	21.7
가을	14.15	0.83	15.5	14.2	12.2

- ③ 차트 메뉴에서 + 버튼을 눌러 시각화할 수 있는 네 가지 형태인 그래프, 막대, 선, 원, 점(산포도) 중에서 한 가지를 선택한다. 이때 어떤 목적으로 그래프를 그릴 것인가에 따라 시각화하는 형태가 달라진다.

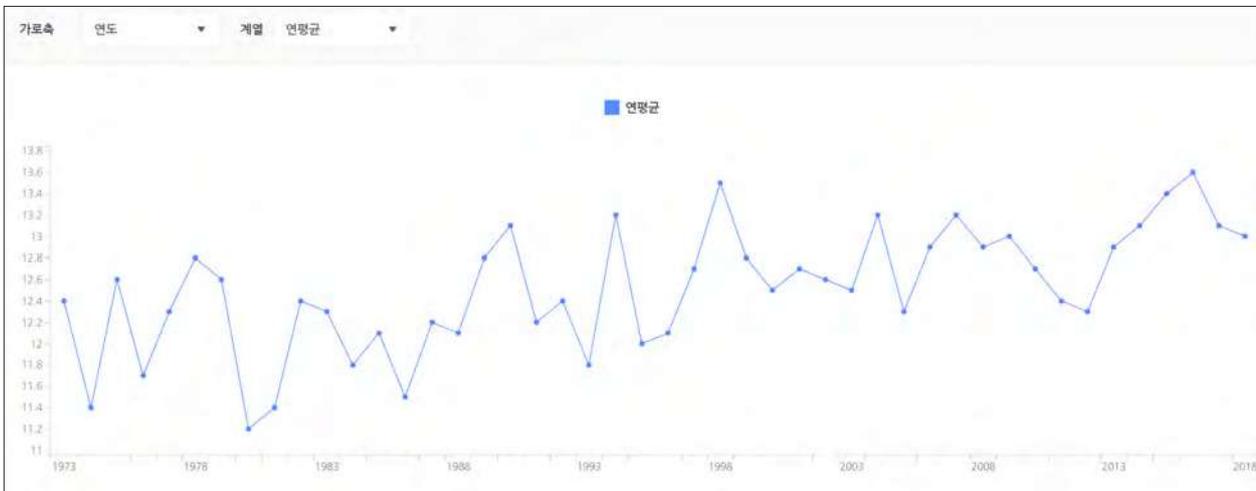
여기서는 계절에 따른 기온 변화 추이를 살펴볼 것이므로 선 그래프를 선택합니다.



④ 차트를 선택하여 차트의 제목을 입력한다.

⑤ 그래프의 가로축(x)과 세로축(y)에 배치할 속성을 선택한다.

예) 연도에 따른 계절별 기온 변화를 살펴볼 때 가로축은 '연도' 속성으로, 계열 (세로축)은 '연평균'으로 설정한다



⑥ 저장하기 버튼을 클릭하여 차트를 저장한다.

저장되었습니다.

테이블과 차트의 정보가 저장되었습니다.

데이터 시각화를 위한
예시므로 분석을 별도로 하지
않으나 점차 연평균 기온이 상승하고
있는 큰 경향성은 알아낼 수
있습니다.



3 문제 해결하기

문제 해결에 필요한 데이터의 속성을 시각화해 보자.

1 데이터 불러오기

플랫폼에서 제공하는 데이터를 불러오기 위해서 **테이블 불러오기** 버튼을 클릭하고, **테이블 추가하기** - '성별/연령별 사용자 분포' 데이터 테이블을 추가한다.

분석에 사용할 데이터는 온라인 사이트나 데이터 제공 사이트, 설문 조사를 통해 수집합니다. 본 단원의 목표가 데이터를 이해하는 것이 핵심이므로 데이터 시각화 플랫폼에서 제공하는 데이터를 사용합니다.



속성은 무엇일까요?

2 데이터 속성 확인하기

1 데이터를 추가하고 '정보' 버튼을 눌러 데이터 속성과 통계값을 확인한다.

요약 네이버 블로그의 각 주제별 조회수 비율을 성별 혹은 연령별로 구분한 값입니다.(%)

테이블 테이블에서 열을 기준으로 한 기초 통계량입니다.

행 22개 | 열 34개 | 셀 748개

	평균	표준 편차	최대값	중간값
연령대	-	-	-	-
성별	-	-	-	-
건강·의학	4.55	3.27	10.35	3.6
게임	4.54	5.31	18.77	2.18
공연·전시	4.54	4.41	14.44	2.66



② 테이블 메뉴를 클릭하여 실제 어떤 데이터가 포함되어 있는지 데이터 속성과 값들을 파악한다. 이때 확인할 수 있는 속성과 값들을 설명해 보자.

네이버 블로그 성별/연령별 사용자 분포

테이블 카드 정보 저장하기

	A	B	C	D	E	F	G
1	연령대	성별	건강·의학	게임	공간·간서	교육·학문	국내여행
2	0~12	여성	0.03	0.1	0.02	0.07	0.02
3	0~12	남성	0.01	0.11	0.01	0.07	0.01
4	13~18	여성	1.74	3.68	1.42	3.87	1.04
5	13~18	남성	0.7	5.98	0.36	1.92	0.64
6	19~24	여성	9.4	7.85	7.38	13.66	6.7
7	19~24	남성	3.6	18.77	2.17	7.25	3.16
8	25~29	여성	9.6	5.28	9.12	8.77	9.56
9	25~29	남성	4.53	18.15	3.25	5.87	4.75
10	30~34	여성	8.88	2.83	10.2	6.78	9.92
11	30~34	남성	4.23	11.85	3.53	4.21	5.28
12	35~39	여성	10.35	2.32	14.43	7.71	12.55

Blank area for notes or observations.

데이터 속성에는 문제 해결에 필요한 것도 있지만 불필요한 것도 있을 수 있어요. 어떤 문제를 해결할 것인가에 따라 필요한 속성을 추출하여 사용합니다.



3 필요한 데이터 추출하기

인터넷 사용 분야에 대한 사용자 분포 데이터 셋을 이용하여 어떤 부분을 살펴볼 것인지에 따라 서로 다른 속성을 추출하여 사용한다. 알고자 하는 정보를 탐색하기 위해 그래프 유형, 가로축과 세로축, 표현 값으로 선택해야 할 속성을 작성해 보자.

알고자 하는 정보	그래프 유형	가로축	세로축	표현 값
성별로 게임에 대해 가장 많이 검색하는 연령대는?				
연령대별로 세계여행에 대한 정보 검색은 여자가 더 많이 할까, 남자가 더 많이 할까?				
국내여행을 검색하는 사람은 세계여행에 대한 검색도 많이 할까? 추이 변화가 남녀 간의 차이가 있을까?				

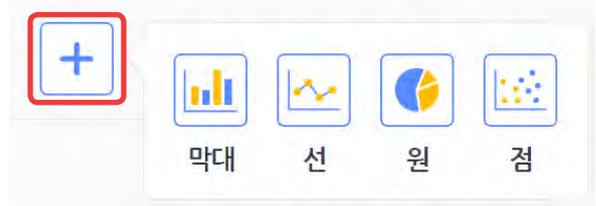
4 데이터 시각화하기

3에서 제시한 세 가지 정보에서 한 가지를 선택하여 파악하기 위한 그래프를 그려 보자.

1 차트 메뉴를 선택한다.

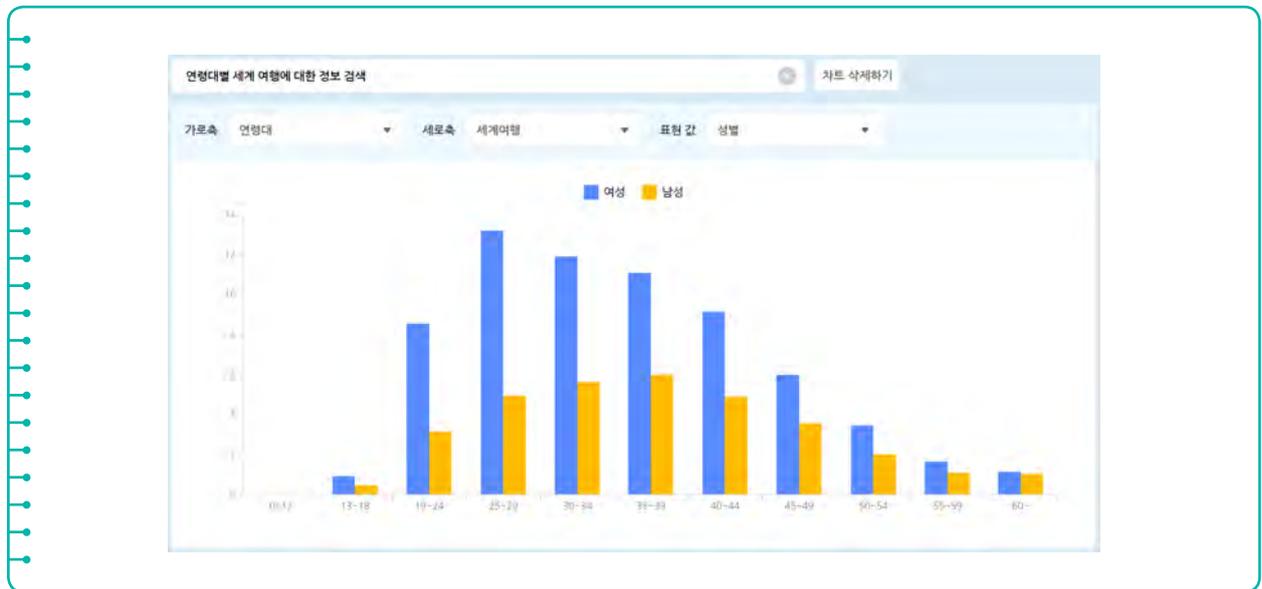


2 + 를 클릭한 후, 알맞은 형태의 그래프를 선택한다.



3 가로축과 세로축, 표현 값을 선택한다.

4 저장하기 버튼을 눌러 그래프를 저장한다.



III 데이터

5 검색 결과 해석하기

결과 그래프는 데이터들을 한눈에 파악하고, 관계를 파악하는 데 도움을 주고 새로운 정보를 제공한다. 결과 그래프를 통해 알 수 있는 정보에서 획득한 정보를 찾아 써 보자.

그래프를 통해 알 수 있는 정보	획득한 정보
성별로 게임에 대해 가장 많이 검색하는 연령대는?	
연령대별로 세계여행에 대한 정보 검색은 여자가 더 많이 할까, 남자가 더 많이 할까?	
국내여행을 검색하는 사람은 세계여행에 대한 검색도 많이 할까? 추이 변화가 남녀 간의 차이가 있을까?	

그래프 이해하기

막대 그래프

막대 그래프는 범주형 데이터를 수직이나 수평으로 표현하여 범주형 데이터의 특징을 살펴볼 때 사용한다.

예) 성별, 연령별 스마트폰의 주당 평균 이용 시간을 비교할 때 막대 그래프를 이용하면 성별 차이와 연령별 차이를 한눈에 파악할 수 있다.

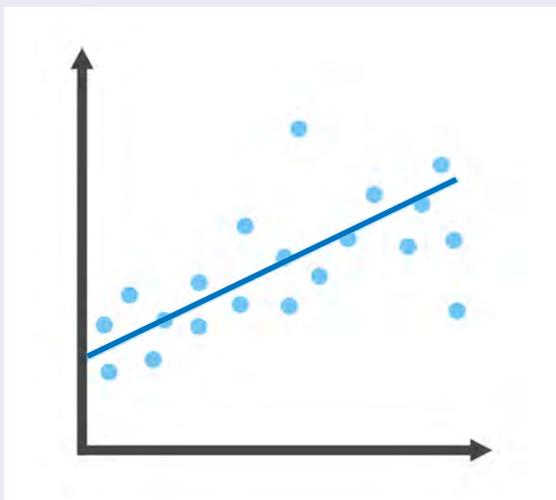


※ 연속적인 자료의 경우 히스토그램을 이용한다.

점 그래프

점 그래프는 데이터의 분포를 살펴보고 변수 간의 관계나 영향력을 볼 때 사용한다. 데이터 분포에서 추세가 존재하면 선을 그려 앞으로의 변화를 예측할 수도 있다.

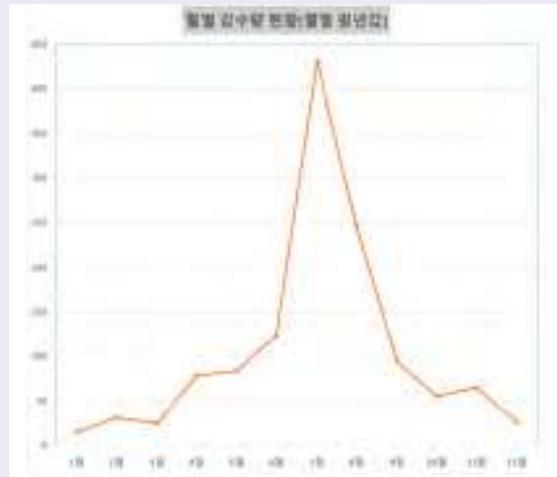
예) 학생들의 학습 시간과 성적 간의 관계와 영향력을 살펴볼 때 점 그래프를 이용하면 쉽게 파악할 수 있다.



선 그래프

선 그래프는 시간에 따라 지속적으로 변화하는 추이를 살펴볼 때 사용한다. 조사하지 않은 시점의 값을 예측할 수 있다는 장점이 있다.

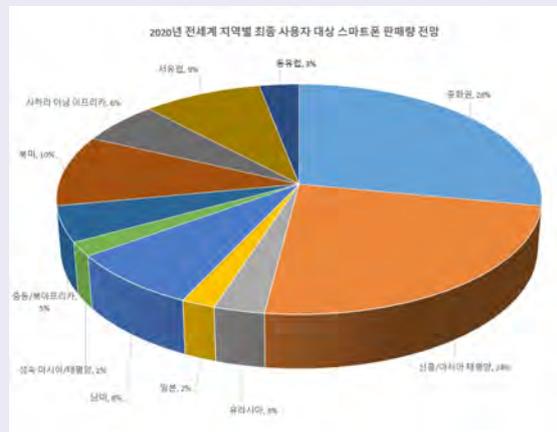
예) 연간 강수량의 변화를 살펴볼 때 선 그래프를 이용하면 월별, 계절별 강수량 변화 추이를 쉽게 파악할 수 있다.



원 그래프

원 그래프는 전체적인 비율을 파악할 때 주로 사용한다. 백분율을 사용할 때 전체에 대한 부분이 차지하는 비율을 한눈에 알 수 있다.

예) 전 세계 지역별 스마트폰의 판매량 비율은 원 그래프를 이용하면 쉽게 파악할 수 있다.



4 정리하기

1 활동을 통해 알게 된 내용을 정리해 보자.

• 데이터 시각화는 _____
 _____ 장점이 있다.

• 그래프의 형태에 알맞은 용도를 적어 보자.

그래프 형태	용도
막대	
선	
원	
점	

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
데이터 시각화 필요성	데이터 시각화를 하는 이유를 설명하였는가?					
그래프 용도	그래프 형태를 목적에 알맞게 사용하였는가?					
참여도	활동에 적극적으로 참여하였는가?					

활동 3

올 겨울 유행할 외투는 무엇일까?

- ▶ 올 겨울 유행할 외투 데이터를 시각화하고 데이터의 의미를 해석하여 새로운 정보를 수집할 수 있다.

1 문제 이해하기

다음 글을 읽고, 포털 사이트의 사용자 검색 데이터를 이용하여 올 겨울 유행할 외투 스타일을 알아보자.

올 겨울에는 어떤 외투가 유행할지 정보를 검색하다 보면 어떤 글에서는 코트가 유행할 것이라고 하고 또 특정 브랜드의 패션 소개를 보면 플리스가 유행할 것이라고 한다.

학교에서 학생들이 입고 다니는 겨울 겉옷으로는 패딩이 여전히 대세인 것도 같다. 어떤 사람들은 무스탕이 패션의 완성이라고도 하며 모피 코트가 정답이라고도 한다. 워낙 다양한 스타일의 옷이 많고 사람들은 각자 좋아하는 스타일의 겉옷을 즐겨 입기도 하니, 올 겨울에는 어떤 스타일의 옷이 유행할지 파악하기 어렵다.

패딩	코트	무스탕
		
플리스	모피 코트	
		

올 겨울에 유행할 외투는 어떤 것일까?



그렇다면, 올 겨울에 많은 사람들이 좋아하고 유행할 옷이 무엇일까?

이번 활동에서 해결해야 할 문제는 무엇인지 써보자.

무엇을 알아야 할까?

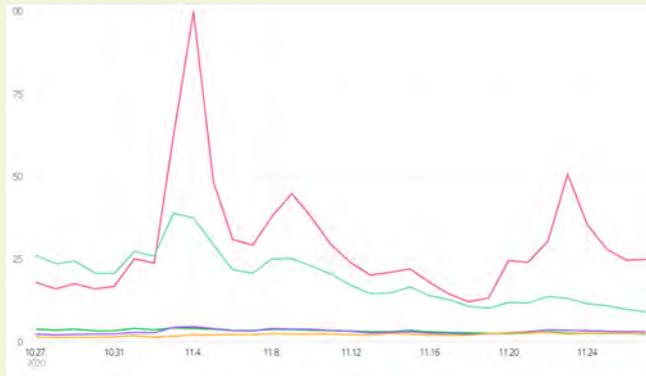
배경 지식

간단한 시각화 도구

데이터를 시각화하는 방법에는 온라인으로 데이터를 불러와 데이터를 분석하는 도구를 사용하기도 하지만, 이번 활동에 사용할 도구처럼 포털 사이트에서 사용자가 입력하는 검색 데이터를 분석하여 검색 트렌드를 파악할 수도 있다.

포털 사이트로 트렌드 파악하기

- 구글의 트렌드나 네이버의 데이터 랩과 같은 포털 사이트에서는 포털 사이트를 이용하는 사용자들이 입력한 검색어 데이터들을 보유하고 있다.
- 검색어 데이터를 사용하면 원하는 데이터를 사람들이 얼마나 검색하는지 기간별 동향을 살펴볼 수 있다.
- 검색 동향에 관한 주제어, 연령대, 성별, 기간, 사용 도구 범위 등을 설정하고 이에 따른 다양한 결과들을 산출할 수 있다.
- 사용하는 전날까지의 검색어를 분석에 사용할 수 있으므로 최신 트렌드를 파악하기에 편리하다. 또한 하루 동안 포털 사이트를 이용하는 사용자의 인원과 입력하는 검색어 수를 고려할 때, 이 데이터들은 매우 방대한 양의 데이터이므로 분석 결과에 대한 신뢰가 매우 높은 편이다.



▲ 날짜별 겨울 외투 검색량 추이

	A	B	C	D	E	F	G	H
1	url	http://datalab.naver.com/keyword/trendResult.naver?hashKey=N_cd47395ace2375						
2	주제	통계						
3	범위	합계						
4	기간	월간 - 2020-08-26 ~ 2020-11-26						
5	성별	전체(여성,남성)						
6	연령대	전체						
7	날짜	패딩	날짜	코트	날짜	플리스	날짜	무스탕
8	2020-08-26	3,11671	2020-08-26	1,38726	2020-08-26	1,68881	2020-08-26	0
9	2020-08-27	2,98482	2020-08-27	1,53483	2020-08-27	1,49847	2020-08-27	0
10	2020-08-28	2,53072	2020-08-28	1,38084	2020-08-28	1,41506	2020-08-28	0
11	2020-08-29	3,11956	2020-08-29	1,73016	2020-08-29	1,47281	2020-08-29	0
12	2020-08-30	3,13382	2020-08-30	1,69451	2020-08-30	1,52413	2020-08-30	0
13	2020-08-31	2,87005	2020-08-31	1,64176	2020-08-31	1,57831	2020-08-31	0
14	2020-09-01	3,111	2020-09-01	1,82853	2020-09-01	1,88556	2020-09-01	0
15	2020-09-02	3,24146	2020-09-02	1,85277	2020-09-02	2,02172	2020-09-02	0
16	2020-09-03	3,3762	2020-09-03	2,17357	2020-09-03	2,23347	2020-09-03	0
17	2020-09-04	3,42253	2020-09-04	2,10941	2020-09-04	2,57777	2020-09-04	0
18	2020-09-05	4,20741	2020-09-05	2,68898	2020-09-05	3,40329	2020-09-05	0
19	2020-09-06	3,19974	2020-09-06	3,51307	2020-09-06	4,561	2020-09-06	0
20	2020-09-07	6,08514	2020-09-07	3,90658	2020-09-07	6,13148	2020-09-07	0
21	2020-09-08	4,59593	2020-09-08	3,31418	2020-09-08	4,35997	2020-09-08	0,00356

▲ 날짜별 겨울 외투 검색 빈도

2 분석 플랫폼 이해하기

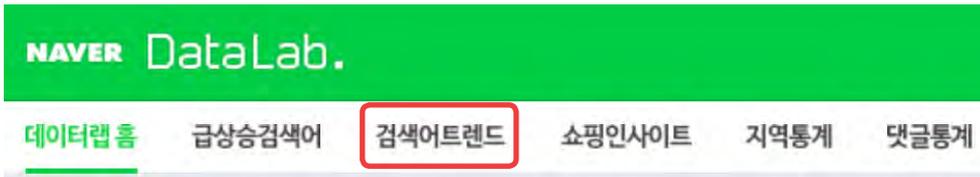
데이터 분석 도구인 데이터 랩(DataLab) 플랫폼 사용 방법을 알아보자.

1 데이터 분석을 위한 데이터 랩 플랫폼 접속하기

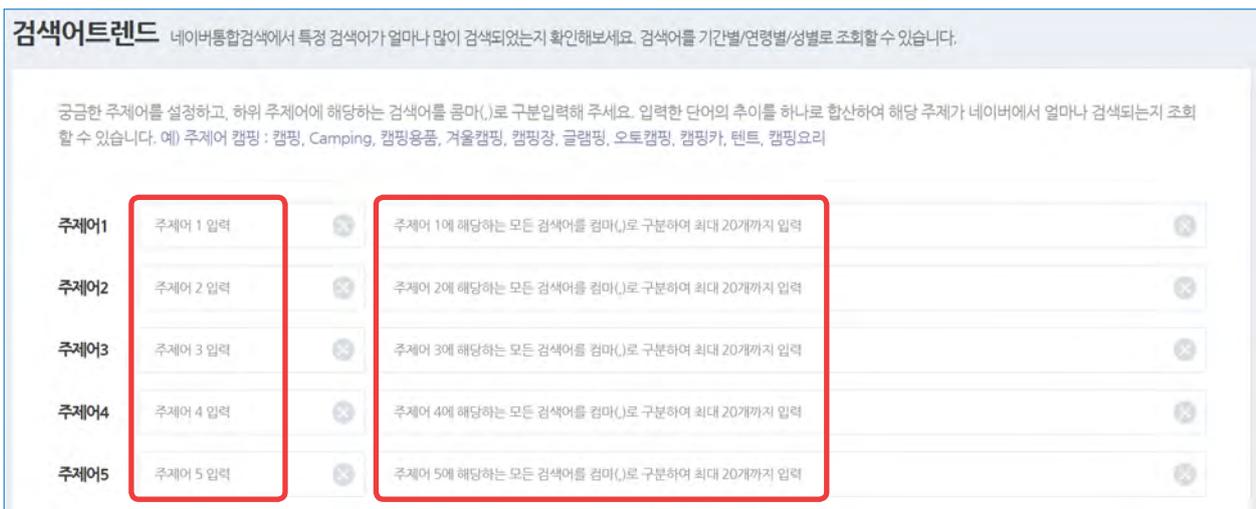
① 주소창에 URL을 입력하여 접속한다.

https://datalab.naver.com/

② 검색 추세를 살펴보기 위해 [검색어트렌드] 메뉴를 선택한다.



③ 비교·분석할 주제어와 주제어를 검색할 때 사용하는 유사 검색어를 입력한다.
유사 검색어가 없으면 입력하지 않아도 된다.



주제어

유사 검색어

입력 방법

궁금한 주제어를 설정하고, 하위 주제어에 해당하는 유사 검색어를 콤마(,)로 구분해서 입력한다.

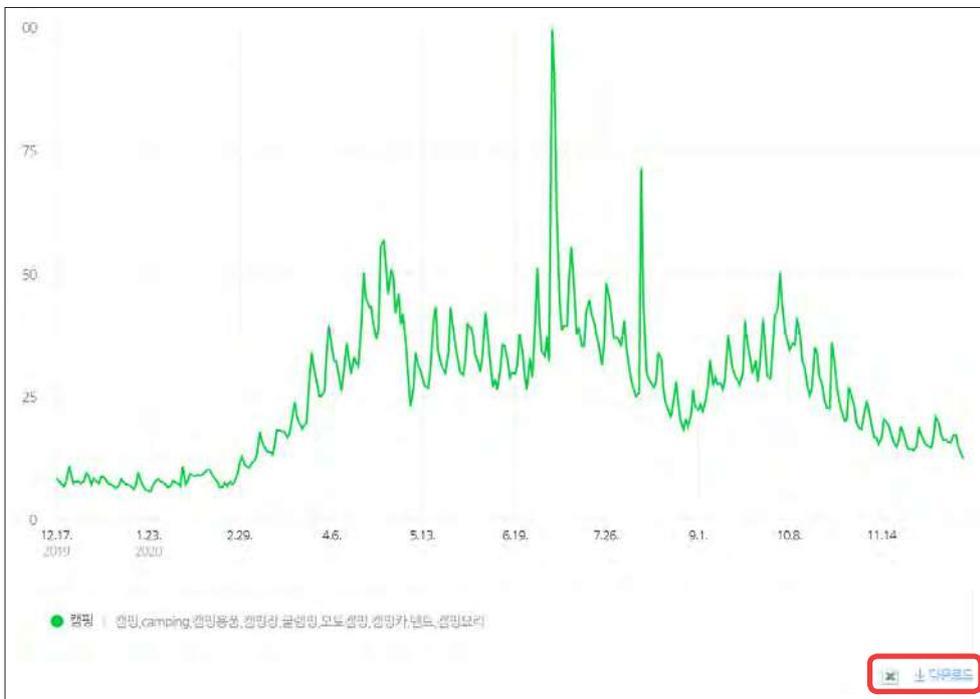
주제어1

사람들이 캠핑에 대해 얼마나 많은 관심을 보이는지 궁금하다면, 주제어1에 '캠핑', 유사 주제어라 할 수 있는 단어인 '캠핑용품', '겨울캠핑', '캠핑장', 'Camping' 등은 주제어 오른쪽 입력란에 작성합니다.



④ 분석하고자 하는 '기간, 범위, 성별, 연령선택'을 체크한다.

⑤  를 클릭하여 결과를 확인한 후 다운로드한다.



III
데
이
터

 **도움 자료**

- 오른쪽 하단의 다운로드를 클릭하면, 그래프를 그리는 데 사용된 데이터를 살펴볼 수 있다.
- 그래프의 수치는 설정한 기간 중 가장 많이 검색된 양을 100으로 설정하여 상대적인 변화가 출력된다.

구글 트렌드(Google Trends) 사용법

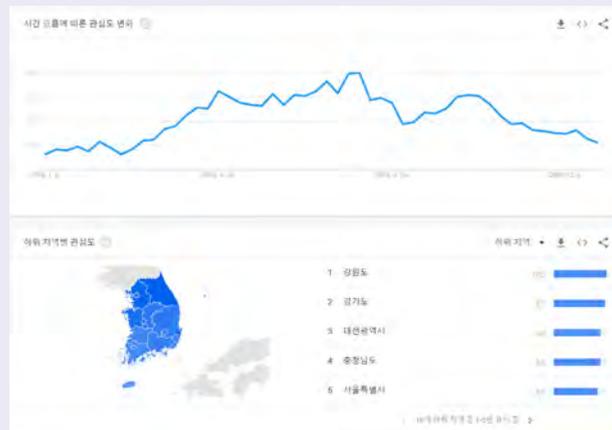
구글 트렌드는 네이버 데이터 랩과 같이 검색 키워드를 입력하여 해당 키워드의 동향을 확인할 수 있다. 관련 주제와 관련 검색어도 함께 확인할 수 있으며, 키워드와 주제를 비교 및 분석도 가능하다.

- 시간 흐름에 따른 관심도 변화: 국가, 기간, 카테고리, 분야 등 선택 가능
- 하위 지역별 관심도: 선택한 국가의 하위 지역의 관심도 제시
- 관련 주제: 인기 주제 또는 인기 급상승하고 있는 관련 주제 제시
- 관련 검색어: 인기 검색어 또는 인기 급상승하고 있는 관련 검색어 제시



구글 트렌드 활용 방법

데이터 랩의 예시와 같이 캠핑에 대한 동향을 확인하는 방법은 다음과 같다.



검색어로 '캠핑'을 입력하면 왼쪽과 같은 결과가 출력되며, 그 결과를 관련 내용에 맞게 해석합니다.



관련 주제	급상승	관련 검색어	급상승
1 저수지 - 주제	급상승	1 감성 캠핑	급상승
2 놀이 - 주제	급상승	2 캠핑 장 노가다	급상승
3 신탄재 - 주제	급상승	3 감성 캠핑	급상승
4 발전기 - 주제	급상승	4 모 등산 캠핑장	급상승
5 베케이션 렌탈 - 숙박 시설 종류	급상승	5 캠핑 제품	급상승

3 문제 해결하기

문제 해결에 필요한 데이터의 속성을 시각화해 보자.

1. 유행할 외투 스타일 찾기

문제 해결에 필요한 데이터를 검색하여 시각화해 보자.

1 검색어 입력하기

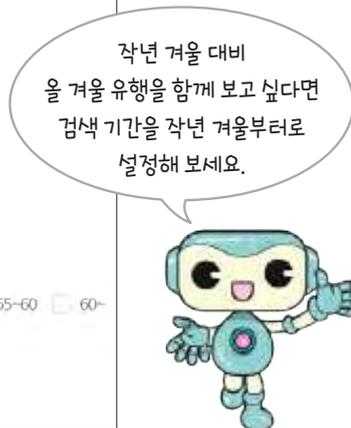
데이터 랩의 [검색어트렌드] 메뉴에서 이번 활동에서 해결하려고 하는 올 겨울 유행하는 외투 동향을 파악하기 위하여 어떤 검색어를 입력해야 하는지 적어 보자.

주제어	유사 검색어



2 검색 상세 설정하기

아래 그림에서 확인하려고 하는 '기간, 범위, 성별, 연령선택'을 체크하고 그렇게 설정한 이유를 적어 보자.



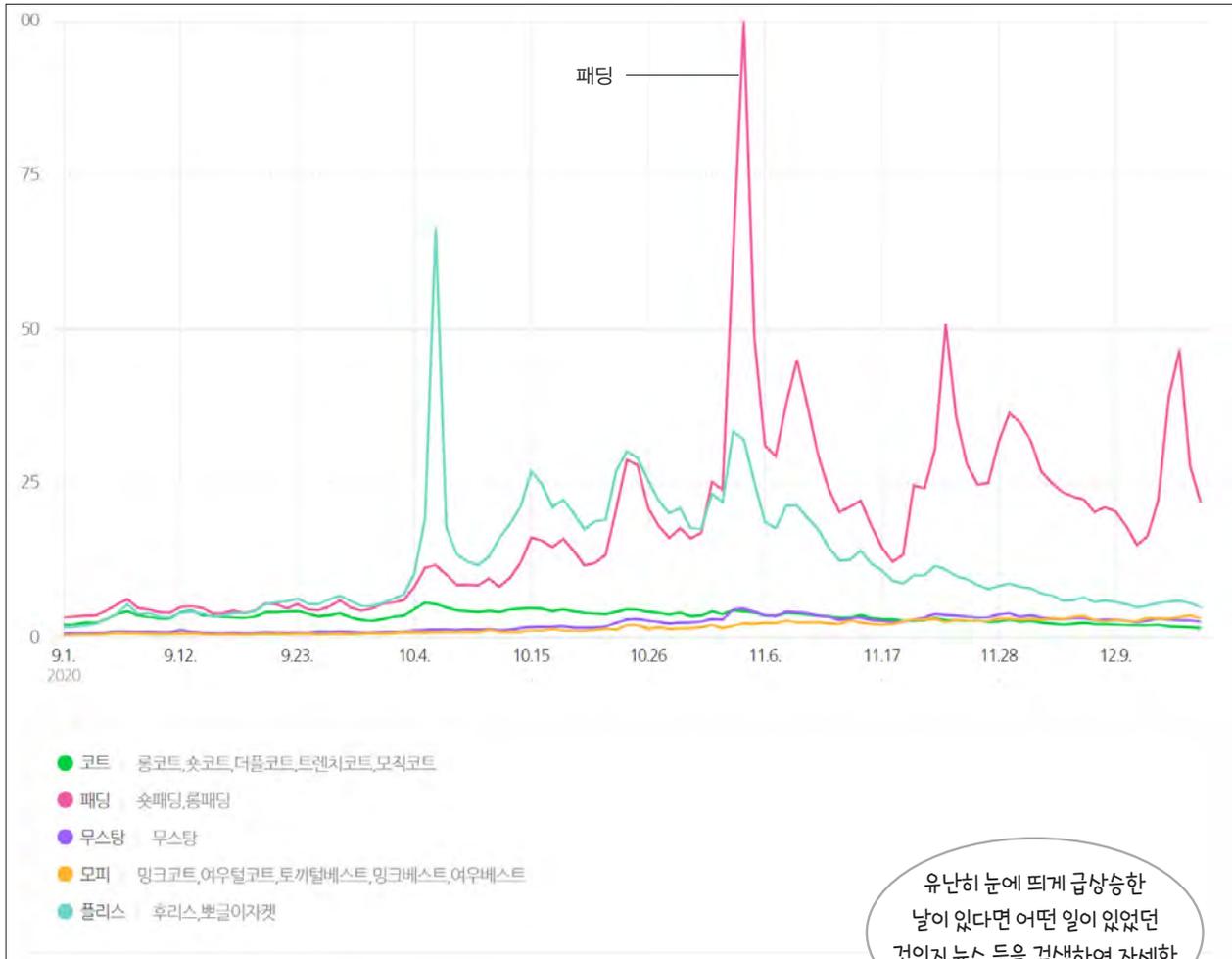
구분	설정된 이유
기간	
범위	
성별	
연령 선택	

3 데이터 시각화하기

뉴스 검색 버튼을 클릭하여 그래프를 출력한다.

4 검색 결과 해석하기

결과 그래프는 데이터들을 한눈에 파악할 수 있고, 관계를 파악하는 데 도움을 주는 새로운 정보를 제공한다. 결과 그래프를 통해 알 수 있는 정보를 찾아보자.



유난히 눈에 띄게 급상승한 날이 있다면 어떤 일이 있었던 것인지 뉴스 등을 검색하여 자세한 정보를 확인해 봅시다!



• 그래프를 통해 알 수 있는 정보를 써 보자.

Blank area for writing information derived from the graph.

4 정리하기

1 활동을 통해 알게 된 내용을 정리해 보자.

• 포털 사이트에서 제공하는 검색어 분석 사이트의 특징은 _____
_____이다.

• 포털 사이트에서 제공하는 검색어 분석 사이트의 장점을 써 보자.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
데이터 분석 도구 활용	문제 해결에 검색어 분석 도구를 사용할 줄 아는가?					
참여도	활동에 적극적으로 참여하였는가?					

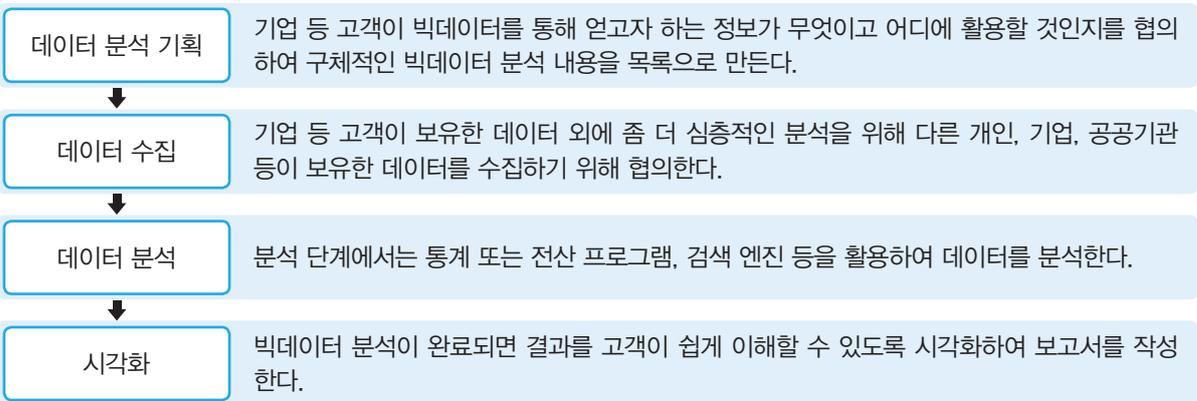
구글의 ‘플루 트렌드’는 감기나 독감의 검색 빈도가 높은 지역을 지도에 표시하여 독감의 확산을 빠르게 예측한다. 페이스북이나 넷플릭스는 개인의 성향과 검색 패턴, 구매 패턴을 분석하여 맞춤형 광고나 영상을 추천한다. 서울시와 KT는 통화 로그 분석을 통해 최적의 심야 시간대 노선을 찾아내기도 했다. 이는 바로 빅데이터가 있기 때문이다.



세상에 존재하는 방대한 양의 데이터를 수집하고 분석하여 가치 있는 정보로 만들어내는 빅데이터는 정치, 경제, 문화, 과학, 교육 등 다양한 영역에서 활용되면서 전 세계가 주목하고 있는 기술이다.

1 빅데이터 전문가는 어떤 일을 할까?

빅데이터 전문가는 거대한 데이터를 실시간으로 수집 및 저장하고, 이 데이터를 분석해 가치 있는 정보를 추출하는 일을 한다. 빅데이터 전문가가 하는 일은 크게 데이터 분석 기획, 데이터 수집, 데이터 분석, 시각화 및 보고서 작성 등으로 구분되며 상세한 내용은 다음과 같다.



2 어떤 분야에서 활동할까?

스마트팩토리 제조 업체, ICT 회사(SI 업체, 포털 사이트, SNS 업체, 통신사 등), 의료 기관, 금융 기관(은행, 증권사, 보험 회사), 게임 회사, 공공 기관, 빅데이터 전문 분석 업체 등

3 어떤 능력과 흥미가 필요할까?

사회 이슈와 트렌드에 대한 이해력, 데이터를 분석할 때의 끈기력, 대용량 데이터 셋(data set)을 코딩하기 위한 프로그래밍 기술, 수치 해석을 위한 수리적 지식, 데이터 분석 도구(각종 통계 패키지) 활용 능력 등

[출처: 4차 산업혁명 시대 내 직업 찾기, 한국고용정보원, 2019]

활동 4

유명 롤러코스터를 살펴볼까?

- ▶ 정형 데이터와 비정형 데이터의 특성을 설명할 수 있다.
- ▶ 유명 롤러코스터 데이터 셋에서 정형 데이터의 종류를 파악하고 시각화하여 문제를 해결할 수 있다.

1 문제 이해하기

다음 글을 읽고, 롤러코스터의 특징을 분석해 보자.

롤러코스터는 설계된 레일 위를 격렬하게 달리도록 만들어진 놀이 기구이다. 전 세계 대부분의 놀이공원에 있으며 놀이공원의 상징으로 여겨지기도 한다. 맨 앞좌석에 타는 것이 시각적인 효과로 가장 무섭다고 하지만 롤러코스터를 즐기는 사람들의 선호도가 높다. 롤러코스터가 중간에 멈추거나 레일을 이탈하지 않도록 하기 위해서는 위치 에너지와 운동 에너지 간의 전환에 관한 전문적인 지식과 정확한 계산을 필요로 한다.

미국은 1915년에서부터 2016년 사이에 157개의 롤러코스터가 제작 및 운영되었다. 어떤 롤러코스터는 나무로 만들어 움직일 때마다 마치 나무가 쓰러질 것 같은 소리 효과를 주기도 하고, 어떤 롤러코스터는 강철로 만들기도 한다. 또 롤러코스터 종류에 따라 앉아서 타는 것과 서서 타는 것으로 구분할 수도 있다. 360도 회전을 하면서 사람이 거꾸로 매달리기도 하는 구간이 있는데, 이 구간의 횟수도 롤러코스터마다 다르다.

이처럼 다양한 롤러코스터는 1915년 이후 어떻게 제작 및 운영되었을까?



이번 활동에서 해결해야 할 문제는 무엇인지 써 보자.

무엇을 알아야 할까?

배경 지식

정형 데이터와 비정형 데이터

데이터는 크게 정형 데이터와 비정형 데이터로 구분할 수 있다.

- 정형 데이터: 의미를 파악하기 쉽도록 규칙에 따라 정리가 된 데이터들을 의미한다.
- 비정형 데이터: 데이터의 의미를 파악하기 쉽지 않고 규칙을 갖지 않는 불규칙한 데이터를 의미한다.

온라인에서 데이터의 약 20%는 정형 데이터에 속하며, 약 80% 정도는 비정형 데이터에 속한다. 정형 데이터의 종류에는 규칙에 따라 정형화된 우편 번호, 주민 등록 번호 등이 있으며, 비정형 데이터 종류에는 이미지, 텍스트, 이메일, 동영상 등이 있다.



▲ 온라인에서 생산되는 데이터

표

Ind. No.	Coaster	Park	City	State	Type	Design	Year	Top Speed (mph)	Max. Height (ft)
1	Ziploc Rigger	Libertyland	Memphis	Tennessee	Wooden	Six Down	1975	42	70
2	Jack Rabbit	Kennywood Park	West Mifflin	Pennsylvania	Wooden	Six Down	1921	45	40
3	Thunderhawk	Dorney Park	Allentown	Pennsylvania	Wooden	Six Down	1921	68	80
4	Clint Digger	Santa Cruz Beach	Santa Cruz	California	Wooden	Six Down	1924	35	70
5	Thunderbolt	Kennywood Park	West Mifflin	Pennsylvania	Wooden	Six Down	1924	56	70
6	Wildcat	Lake Compounce	Bristol	Connecticut	Wooden	Six Down	1927	48	85
7	Coaster Thrill Ride	Puyallup Park	Puyallup	Washington	Wooden	Six Down	1935	50	55
8	Cyclone	Lakeside	Denver	Colorado	Wooden	Six Down	1940	50	90
9	Comet	Hersheypark	Hershey	Pennsylvania	Wooden	Six Down	1940	50	90

▲ 롤러코스터에 대한 정형 데이터 예

텍스트

<롤러코스터 이야기>

미국은 1915년에서부터 2016년 사이에 157개의 롤러코스터가 제작 및 운영되었다. 어떤 롤러코스터는 나무로 만들어 움직일 때마다 마치 나무가 쓰러질 것 같은 소리 효과를 주기도 하고, 어떤 롤러코스터는 강철로 만들기도 한다.

▲ 롤러코스터에 대한 비정형 데이터 예

2 분석 플랫폼 이해하기

본 활동에서는 데이터 분석을 위하여 코답(CODAP) 플랫폼을 이용한다.

1 코답(CODAP) 플랫폼 접속하기

- 주소창에 URL을 입력하여 접속한다. 이때크롬() 브라우저에서 접속해야 오류 없이 사용할 수 있다.

`https://codap.concord.org/`

2 데이터가 준비되어 있는 경우

사전에 데이터 파일을 가지고 있다면 분석만 실시하면 되므로, 사이트의 우측 상단의  버튼을 이용하여 프로그램을 실행한다.

코답(CODAP)은 영문으로 된 플랫폼이지만, 크롬에서 마우스 오른쪽 버튼 클릭하여 '한국어로 번역'으로 변환하면 사용하기 편리합니다.



3 코답에서 제공하는 예시 데이터를 사용하는 경우

프로그램 실행 전 예시 데이터를 선택하고 분석하는 것이 편리하다. 따라서 사이트 중간 위치의 'Sample CODAP Datasets'을 이용하는 것이 데이터를 선택하는 데 편리하다.



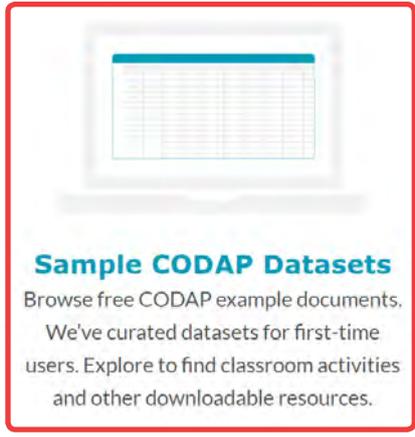
Try CODAP

CODAP is easy to use and runs in your web browser. CODAP is (and always will be) free. Share your data with others and bring it to life!



Get the Code

Learn how to create your own plugin in CODAP. Explore our open-source GitHub repository and start incorporating it into your work today.



Sample CODAP Datasets

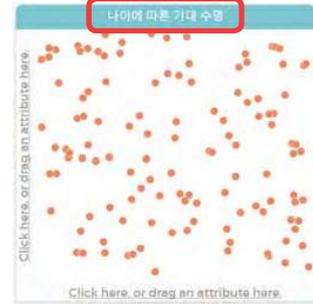
Browse free CODAP example documents. We've curated datasets for first-time users. Explore to find classroom activities and other downloadable resources.

2 데이터 시각화 방법

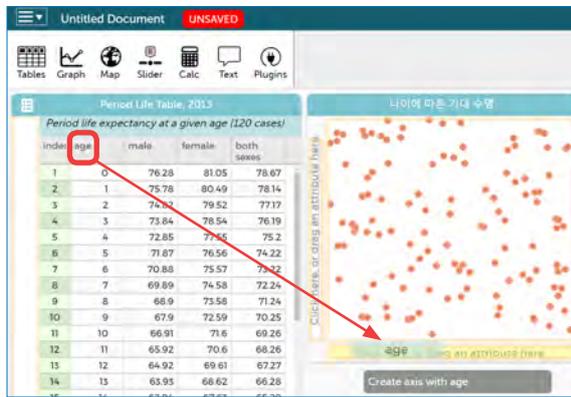
① 상단의 메뉴 중에서 그래프(Graph)를 선택한다.



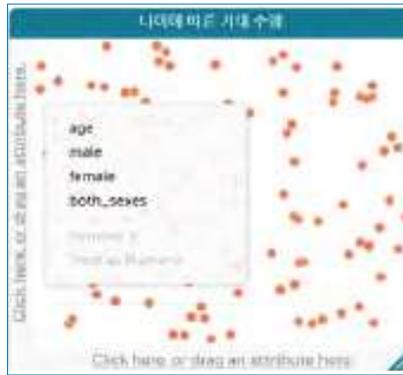
② 임의의 산점도가 생기면 차트 이름을 '나이에 따른 기대 수명'으로 변경한다.



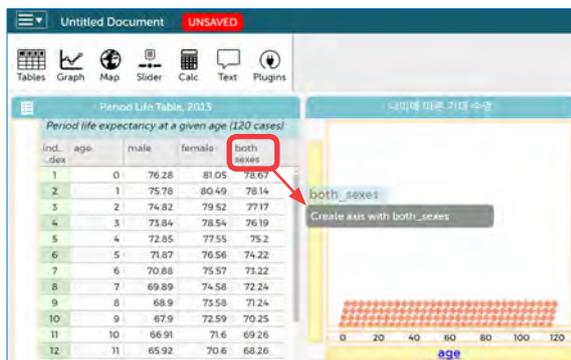
③ 테이블의 나이(age) 속성을 그래프의 x축으로 드래그 앤 드롭(drag&drop)하여 x축을 생성한다.



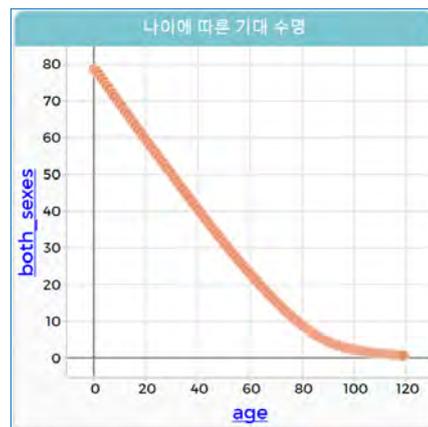
그래프의 x축, y축의 'click here' 부분을 클릭하여 속성을 추가할 수 있습니다.



④ 테이블의 남녀 합한 기대 수명(both sexes) 속성을 그래프의 y축으로 드래그 앤 드롭(drag&drop)하여 y축을 생성한다.



⑤ 사람들이 해당 나이에서 얼마나 더 살 수 있을지 기대하는 그래프가 완성된다.



3 정형 데이터와 비정형 데이터 살펴보기

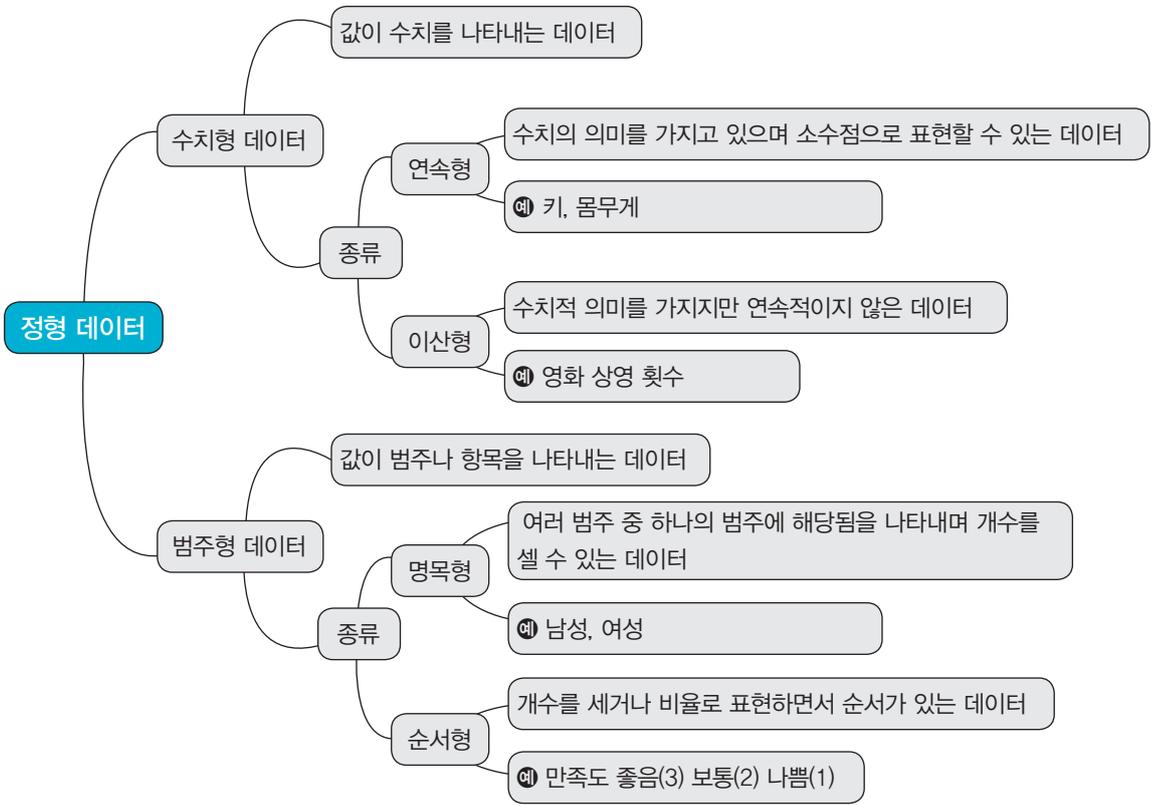
롤러코스터에 대한 비정형 데이터를 정형 데이터로 만든 롤러코스터 데이터 셋을 통해 데이터의 속성과 속성별 데이터의 유형을 자세히 살펴보자.

데이터는 규칙이나 형식이 정해져 있느냐에 따라 정형 데이터와 비정형 데이터로 나눌 수 있다. 속성을 통해 시각화, 관계 분석 등이 용이하므로 정형 데이터를 선호한다. 하지만, 비정형 데이터는 전체의 80%를 차지하므로 데이터 분석에 활용하기 위해서 정형 데이터로 변환해서 데이터 분석에 사용된다.

구분	정형 데이터	비정형 데이터
예시		<p>〈롤러코스터 이야기〉</p> <p>미국은 1915년에서부터 2016년 사이에 157개의 롤러코스터가 제작 및 운영되었다. 어떤 롤러코스터는 나무로 만들어 움직일 때마다 마치 나무가 쓰러질 것 같은 소리 효과를 주기도 하고, 어떤 롤러코스터는 강철로 만들기도 한다.</p>
특징	의미를 파악하기 쉽도록 규칙에 따라 정리되어 있다.	데이터의 의미를 파악하기 쉽지 않고 규칙을 갖지 않은 불규칙성을 띤다.

1 정형 데이터의 종류 살펴보기

정형 데이터에는 크게 수치형 데이터와 범주형 데이터로 구분하며, 수치형 데이터와 범주형 데이터는 각각 아래 그림과 같이 표현할 수 있다.



2 롤러코스터 데이터 셋의 속성 살펴보기

롤러코스터 데이터 셋에는 아래 그림에 표시된 속성을 갖고 있다.

ind. _dex	Age Group	ind. _dex	Coaster	Park	City	State	Type	Design	Year Opened	Top Speed (mph)	Max Height (ft)	Drop (ft)	Length (ft)	Duration (s)	Inversions	Num of versions
1	1.older	1	Zippin ...	Liberty...	Memphis	Tenness...	Wooden	Sit Down	1915	40	70	70	2865	90	N	0
2	2.rec...	2	Jack Ra...	Kennyw...	West Mi...	Pennsyl...	Wooden	Sit Down	1921	45	40	70	2132	96	N	0
3	3.new...	3	Thunde...	Dorney ...	Allento...	Pennsyl...	Wooden	Sit Down	1923	45	80	65	2767	78	N	0
4		4	Giant Di...	Santa C...	Santa C...	Californ...	Wooden	Sit Down	1924	55	70	65	2640	112	N	0
5		5	Thunde...	Kennyw...	West Mi...	Pennsyl...	Wooden	Sit Down	1924	55	70	95	2887	101	N	0
6		6	Wildcat	Lake Co...	Bristol	Connec...	Wooden	Sit Down	1927	48	85	78	2746	75	N	0
7		7	Coaster ...	Puyallu...	Puyallup	Washin...	Wooden	Sit Down	1935	50	55	52	2650	105	N	0
8		8	Cyclone	Lakeside	Denver	Colorado	Wooden	Sit Down	1940		90	89	2800	120	N	0

도움 자료

각 속성별 의미

롤러코스터의 이름(Coaster), 설치된 공원(Park), 도시(City), 주(State), 만든 재료(Type), 탑승 방식(Design), 개시 연도(Year), 최고 속도(Top Speed), 최고 높이(Max Height), 최고 높이와 최저 높이의 차이(Drop), 트랙의 길이(Length), 탑승 시간(Duration), 뒤집기 기능(Inversions), 뒤집기 횟수(Num of versions), 제작 시기 그룹(Age Group) 속성이다.

3 정형 데이터 종류 구분하기

앞의 정형 데이터 예시를 참고하여 롤러코스터 데이터 셋의 속성들을 나누어 보자.

속성	정형 데이터 종류	속성	정형 데이터 종류
롤러코스터의 이름(Coaster)		탑승 방식(Design)	
설치된 공원(Park)		트랙의 길이(Length)	
도시(City)		탑승 시간(Duration)	
주(State)		뒤집기 횟수(Num of versions)	
만든 재료(Type)		제작 시기 그룹(Age Group)	

4 문제 해결하기

문제 해결에 필요한 데이터의 속성을 찾아 시각화해 보자.

1 데이터 불러오기

플랫폼에서 제공하는 데이터를 불러오기 위해서 'Sample CODAP Datasets'을 클릭하고, 'Roller Coasters'를 선택한다.



2 데이터 속성 확인하기

수집한 데이터를 불러와 생성된 테이블에 포함되어 있는 속성을 확인한다. 데이터 속성에는 문제 해결에 필요한 것도 있지만 불필요한 것도 있으므로 어떤 문제를 해결할 것인가에 따라 필요한 속성을 추출하여 사용한다.

ind. _dex	Coaster	Park	City	State	Type	Design	Year Opened	Top Speed (mph)	Max Height (ft)	Drop (ft)	Length (ft)	Duration (s)	Inversions	Num of _versions
1	Zippin	Liberty	Memphis	Tenness	Wooden	Sit Down	1915	40	70	70	2865	90	N	0
2	Jack Ra	Kennyw	West Mi	Pennsyl	Wooden	Sit Down	1921	45	40	70	2132	96	N	0
3	Thunde	Dorney	Allento	Pennsyl	Wooden	Sit Down	1923	45	80	65	2767	78	N	0
4	Giant Di	Santa C	Santa C	Californ	Wooden	Sit Down	1924	55	70	65	2640	112	N	0
5	Thunde	Kennyw	West Mi	Pennsyl	Wooden	Sit Down	1924	55	70	95	2887	101	N	0
6	Wildcat	Lake Co	Bristol	Connec	Wooden	Sit Down	1927	48	85	78	2746	75	N	0
7	Coaster	Puyallu	Puyallup	Washin	Wooden	Sit Down	1935	50	55	52	2650	105	N	0
8	Cyclone	Lakeside	Denver	Colorado	Wooden	Sit Down	1940		90	89	2800	120	N	0
9	Comet	Hershey	Hershey	Pennsyl	Wooden	Sit Down	1946	50	84	78	3360	105	N	0
10	Comet	Walda	Erie	Pennsyl	Wooden	Sit Down	1951	25	37	25	1300	84	N	0
11	High Sp	Knoebels	Elysburg	Pennsyl	Steel	Sit Down	1955	40	18	12	200		N	0
12	Bobsleds	Seabree	Rochest	New Yo	Steel	Sit Down	1962	23	31	16	1240		N	0

3 문제 해결을 위해 필요한 속성 추출하고 시각화하기

다음 문제를 해결하기 위해 필요한 속성을 찾아 시각화하고 그 의미를 해석해 보자.

① 가장 높은 롤러코스터와 가장 빠른 롤러코스터의 이름 및 높이, 속도를 찾아보자.

- 필요한 속성을 찾아보자.

- 데이터를 시각화해 보자.

해당 점 위에 마우스를 올려 놓으면 롤러코스터의 이름(고유한 속성)과 높이 또는 속도를 확인할 수 있다.

- 시각화한 결과를 토대로 문제의 답을 찾아보자.

② 최근에 만들어진 롤러코스터와 오래된 롤러코스터 중에서 어느 롤러코스터가 더 빠른지 알아보자.

- 필요한 속성을 찾아보자.

Empty rounded rectangular box for writing.

- 데이터를 시각화해 보자.

Large empty rounded rectangular box for drawing a visualization.

- 시각화한 결과를 토대로 문제의 답을 찾아보자.

Empty rounded rectangular box for writing the answer.

③ 제작된 시기에 따라 사용된 유형(나무, 강철)의 변화를 알아보자.

- 필요한 속성을 찾아보자.

Empty rounded rectangular box for writing.

- 데이터를 시각화해 보자.

Large empty rounded rectangular box for drawing a visualization.

- 시각화한 결과를 토대로 문제의 답을 찾아보자.



④ 제작된 시기에 따라 롤러코스터 길이의 변화를 알아보자.

- 필요한 속성을 찾아보자.



- 데이터를 시각화해 보자.



- 시각화한 결과를 토대로 문제의 답을 찾아보자.



5 정리하기

1 활동을 통해 알게 된 내용을 정리해 보자.

• 정형 데이터는 _____ 이고,
비정형 데이터는 _____ 이다.

• 정형 데이터의 유형으로 _____
_____ (이)가 있다.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
정형 데이터와 비정형 데이터의 정의	정형 데이터와 비정형 데이터가 무엇인지 이해하였는가?					
정형 데이터의 종류	정형 데이터의 종류를 이해하고 특성에 맞게 사용할 수 있는가?					
참여도	활동에 적극적으로 참여하였는가?					

활동 5

남녀공학 여부에 따른 나의 성적 평가는?

- ▶ 문제 해결에 필요한 데이터를 수집하여 분석할 수 있다.
- ▶ 남녀공학 여부에 따른 학생들의 성적에 대한 평가를 파악하기 위해 필요한 속성을 가공하여 데이터를 시각화한 후 분석할 수 있다.

1 문제 이해하기

다음 글을 읽고, 남녀공학 여부에 따른 학생들의 성적에 대한 평가를 파악하기 위하여 필요한 데이터를 찾아 그 관계를 분석해 보자.

과거에는 남학교와 여학교를 명확하게 구분했지만, 어느 때부터인가 남학생과 여학생을 한 학교에 다니게 하는 남녀 공학이 생기기 시작했다. 그리고 남녀 공학이라고 하더라도 남학생과 여학생 합반인 학교가 있고 분반인 학교가 있다.

어떤 사람들은 함께 하는 사회에서 서로를 이해하고 함께 협업하는 역량도 길러야 하니 남녀 공학을 다닐 필요가 있다고 하고, 또 다른 사람들은 이성에 관심을 가지게 되면서 학업에 소홀할 수 있다고 이야기한다.

어떤 기사들은 남녀공학에서 우수한 성적을 거두는 학생들이 가장 많이 배출된다고도 한다.

남·여·공학 고교 유형별
2017·2018학년도 수능 표준점수 평균

영역	학교 유형	2017학년도	2018학년도
국어	남고	99.4	98.2
	여고	101.9	101.6
	공학	96.3	96.0
	전체	98.3	97.8
수학가	남고	101.2	100.3
	여고	100.4	99.6
	공학	96.8	96.0
	전체	98.9	98.0
수학나	남고	100.5	100.4
	여고	101.4	101.0
	공학	97.2	97.4
	전체	99.0	98.9
영어	남고	99.2	
	여고	101.0	
	공학	96.3	절대평가 시행
	전체	98.1	

[출처: 한국교육과정 평가원]



그렇다면 학생들은 남녀공학에서는 자신이 받은 성적에 대해 어떤 평가를 할까?

이번 활동에서 해결해야 할 문제는 무엇인지 써 보자.

.....

.....

무엇을 알아야 할까?

배경 지식

데이터 수집 및 가공 방법

인공지능은 문제를 해결하기 위하여 방대한 양의 데이터를 이용한다. 데이터를 수집하는 방법을 살펴보고 문제 해결에 필요한 데이터를 수집하여 시각화한 후 분석하는 가공 방법을 알아보자.

공공 데이터 검색	민간 데이터 검색	직접 수집
<ul style="list-style-type: none"> • 정부나 지방자치단체 등의 기관이 데이터를 수집, 생성, 관리 • 교육, 사회 복지, 문화 관광 등 다양한 데이터 포함. • 무료로 사용 가능 	<ul style="list-style-type: none"> • 기업이 데이터를 수집, 생성, 관리 • 다양한 주제 데이터 사용 가능 • 주로 수익을 위해 제공함. 	<ul style="list-style-type: none"> • 설문 조사 등을 통한 필요한 데이터 직접 수집 • 원하는 데이터 수집 가능 • 오랜 시간과 비용 소요



▲ 공공 데이터

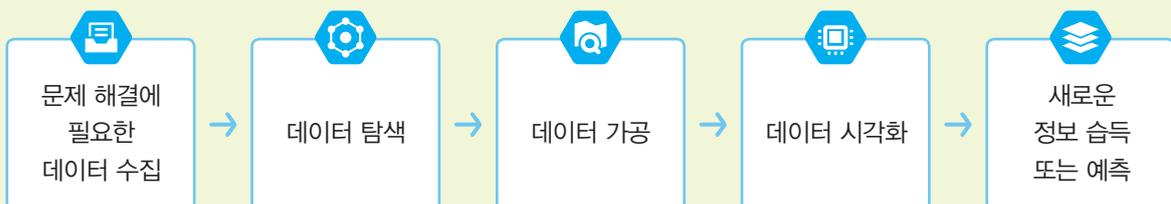


▲ 민간 데이터



▲ 직접 수집한 데이터

수집한 데이터 셋에는 문제 해결에 필요한 데이터만 있는 것이 아니므로, 데이터로부터 필요한 데이터를 가공할 필요가 있다. 이때 데이터의 개수를 세거나 평균 또는 표준편차와 같은 통계적 수치를 사용해야 할 때는 데이터 값으로부터 새로운 데이터를 가공해야 한다. 또한 데이터를 처리할 수 있는 형태로 수정해야 하는데, 값이 없는 데이터나 이상한 값을 가진 데이터 등이 결과에 영향을 미치지 않도록 상황에 따라 데이터를 삭제하거나 평균값으로 대체하는 등의 방식으로 가공한다.



2 분석 플랫폼 이해하기

데이터 분석 도구인 코답(CODAP) 플랫폼 사용 방법을 알아보자.

1 데이터 분석을 위한 코답(CODAP) 플랫폼 접속하기

① 주소창에 URL을 입력하여 접속한다.

`https://codap.concord.org/`

② 데이터가 준비되어 있는 경우

사전에 데이터 파일을 가지고 있다면 분석만 실시하면 되므로, 사이트의 우측 상단의

Try CODAP 버튼을 이용하여 프로그램을 실행한다.

③ 예시 데이터를 사용할 경우

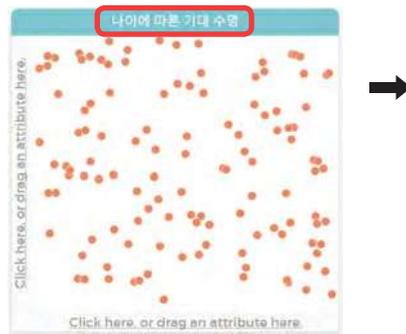
코답 사이트에서 'Sample CODAP Datasets'을 이용한다.

2 데이터 시각화하기

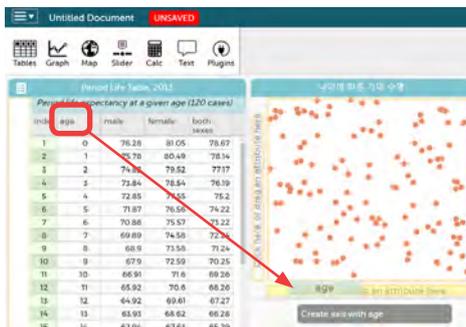
① 상단 메뉴에서 그래프를 선택한다.



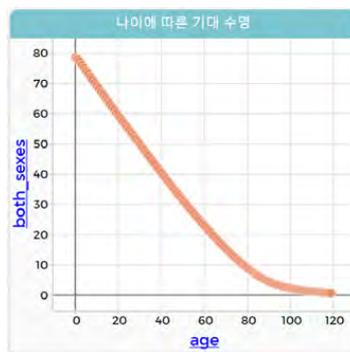
② 임의의 산점도가 생기면 차트 이름을 정한다.



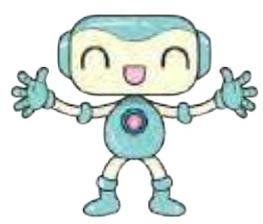
③ 속성을 정하여 x축과 y축을 생성한다.



④ 해당하는 그래프가 완성된다.



분석 플랫폼 코답 사이트 연결 과정은 앞 활동을 참고하세요.



3 문제 해결하기

문제 해결에 필요한 데이터를 수집해 보자.

1. 남녀공학 구분에 따른 전체 성적 주관적 평가 차이 분석하기

‘남녀공학 여부에 따른 나의 성적 평가’를 확인하기 위해 관련 데이터를 제공하는 사이트를 탐색해 보자.

1 데이터 수집하기

① 주소창에 URL을 입력하여 접속한다.

```
https://www.nypi.re.kr/archive
```

한국 아동 청소년 데이터 아카이브
아동·청소년, 다문화 청소년 패널 조사 등 각종 통계 자료를 무료로 제공합니다.



III 데이터

다음 자료

한국 아동·청소년 데이터 아카이브(<https://www.nypi.re.kr/archive>)

한국청소년정책연구원은 수집한 데이터를 활용하여 아동·청소년의 의식이나 생활을 더욱 체계적으로 이해하기 위하여 한국 아동·청소년 데이터 아카이브를 구축하고 있다. 사이트에는 많은 사람들이 연구에 이용할 수 있도록 그동안 수집한 다양한 연구 데이터를 제공하고 있다.

② 화면 상단 메뉴에서 ‘데이터 다운로드’의 ‘조사표/데이터/코드북’을 선택한다.



웹 사이트 업데이트 시점에 따른 화면 구성이 현재와 다를 수 있습니다.



③ 동의 사항에 체크한 다음 아래의 해당 항목을 선택하고 '확인' 버튼을 누릅니다.

○ 데이터신청 동의

조사표/데이터/코드북의 데이터 다운로드를 원하시면 아래의 사항에 동의하여 주십시오.

- 한국청소년정책연구원에서 생산한 데이터의 양도 및 대여를 금지합니다. 또한 자료의 직접 신청자가 아닌 경우는 논문이나 기타 연구보고서에 한국청소년정책연구원에서 생산한 데이터를 이용한 연구결과를 게재하거나 출판할 수 없습니다.
- 한국청소년정책연구원에서 생산한 자료에 대한 저작권과 지적 소유권은 한국청소년정책연구원이 가지고 있으며, 이 권리는 대한민국의 저작권법과 국제저작권 조약에 의하여 보호됩니다. 따라서 사용자는 본 자료를 무단으로 복제하여 판매 또는 불특정 다수에게 배포할 수 없습니다.
- 사용자는 논문이나 기타 연구보고서에 한국청소년정책연구원에서 생산한 데이터를 활용한 연구결과를 게재하거나 출판할 때, 한국청소년정책연구원 데이터를 활용한 것임을 명기해 주셔야 합니다. 또한 게재 및 출판 결과를 한국청소년정책연구원 담당자(nyplidata@nypl.re.kr)에게 출판 서지사항을 알려주시면 감사하겠습니다.

동의합니다. 동의하지 않습니다.

④ '조사선택' 항목은 '패널조사', 하위 분류는 '한국청소년패널조사 2010'을 선택하여 검색한다.

조사선택

조사선택

패널조사

횡단조사

분류선택

한국청소년패널조사

한국아동·청소년패널조사 2010

한국아동·청소년패널조사 2018

다문화청소년패널조사

학업중단청소년패널조사

코드북은 데이터셋에서 알파벳과 숫자로 구성된 데이터 속성이 무엇을 나타내는지 알려 줍니다.



⑤ '[데이터]KCYPs 2010, 중1 코호트 제1-7차 조사 데이터'와 '[코드북]KCYPs 2010 제1-7차 조사 코드북'을 선택한다.

패널조사 - 한국아동·청소년패널조사 2010 - 제목 - 검색

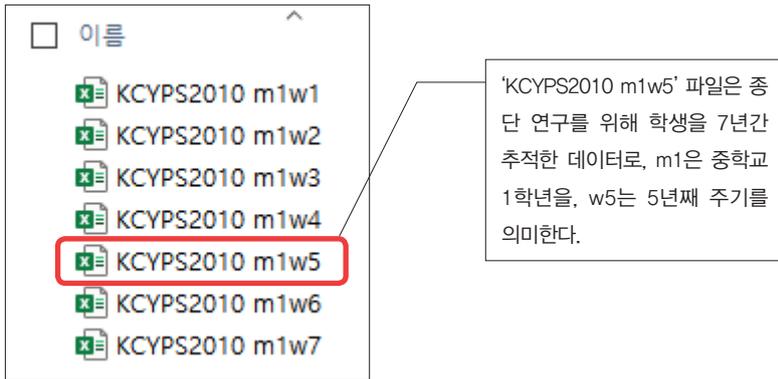
전체 6건, 현재 페이지 1/1

번호	조사유형	자료목록	제목	등록일	조회수	첨부
6	패널조사	한국아동·청소년패널조사 2010	[데이터] KCYPs 2010 중1 코호트 제1-7차 조사 데이터	2015-05-21	15,667	כן
5	패널조사	한국아동·청소년패널조사 2010	[데이터] KCYPs 2010 초4 코호트 제1-7차 조사 데이터	2015-05-21	8,995	כן
4	패널조사	한국아동·청소년패널조사 2010	[코드북] KCYPs 2010 제1-7차 조사 코드북	2015-05-21	21,141	כן
3	패널조사	한국아동·청소년패널조사 2010	[조사표] KCYPs 2010 제1-7차 조사 조사표	2015-05-21	14,150	כן
2	패널조사	한국아동·청소년패널조사 2010	[유저가이드] KCYPs 2010 제1-7차 조사 유저가이드	2015-05-21	11,077	כן
1	패널조사	한국아동·청소년패널조사 2010	[데이터] KCYPs 2010 초1 코호트 제1-7차 조사 데이터	2015-05-21	10,906	כן

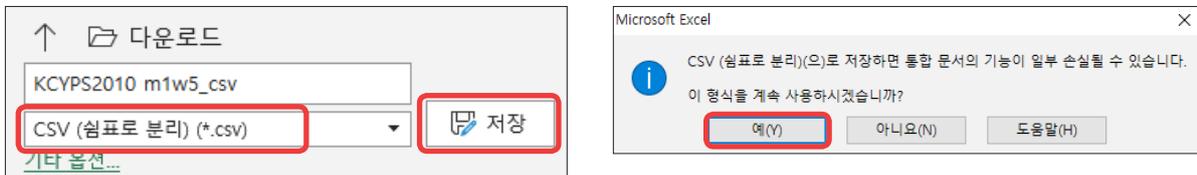
⑥ 첨부 파일의 데이터 중 'EXCEL' 데이터를 선택하여 다운로드한다.



⑦ 데이터 파일의 압축을 풀고, 'KCYP2010 m1w5' 파일을 연다.

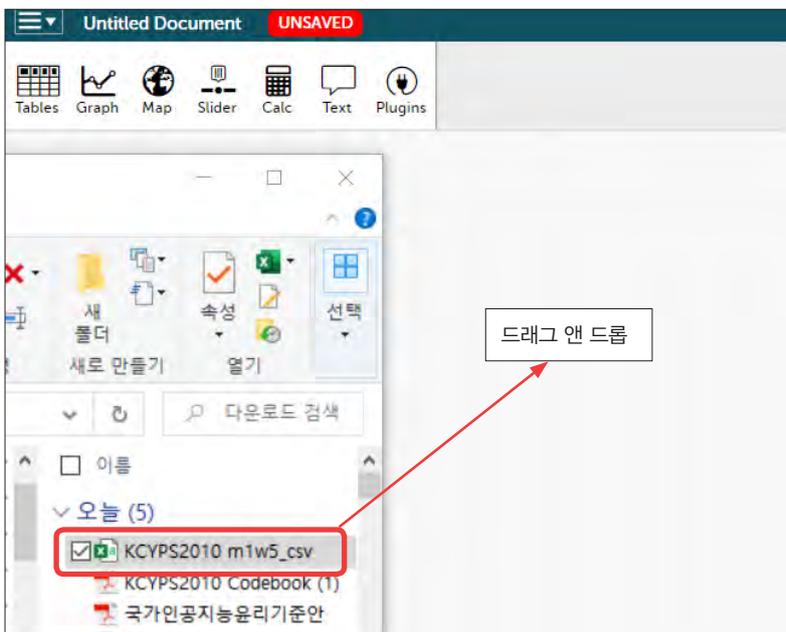


⑧ 코답은 csv(comma-separated values) 형식만 지원하므로 [파일] - [다른 이름으로 저장]에서 파일 형식을 CSV로 선택하고 다시 저장한다.



2 데이터 불러오기

① CSV 파일을 코답 사이트로 직접 드래그 앤 드롭하여 불러온다.



② 데이터 테이블과 속성을 확인한다.

Ind... ...dex	ID	SCLIDw5	PANEL... ...NELw5	WAVEw5	SURVEY... ...VEY1w5	SURVEY... ...VEY2w5	WEIGHT... ...GHT1w5	WEIGHT... ...HT2w5	GI
1994	148922	1981	3	5	1	1	205.66	304.86	
1995	148923	1978	3	5	1	1	205.66	177.61	
1996	148924	1976	3	5	1	1	205.66	171.27	
1997	148925	1952	3	5	1	1	205.66	171.12	
1998	148926	1986	3	5	1	1	205.66	168.82	
1999	148927	1975	3	5	1	1	205.66	180.46	
2000	148928	1979	3	5	1	1	205.66	175.1	

데이터의 용량이 클 경우에는 약 200개씩 순차적으로 불러오므로 전체 데이터를 업로드하는 데 시간이 오래 걸려요.



③ 코드북을 열어 각 속성이 의미하는 바가 무엇인지 확인한다.

- 코드북의 75쪽에서 '남녀공학 구분' 속성을, 79쪽에서 '전체 성적 주관적 평가' 속성을 찾는다.

변인명	문항번호							변인 설명	변인값
	제1차	제2차	제3차	제4차	제5차	제6차	제7차		
SCHOOL	■	■	■	■	■	■	—	학교 일반(0)	급 1 초등학교 2 중학교 3 고등학교 4 대학교 5 국제학교 6 대안(학교) 7 기타학교
GRADE	■	■	■	■	■	■	—	학년	1 1학년 2 2학년 3 3학년 4 4학년 5 5학년 6 6학년
COEDU	■	■	■	Y03	Y02	■	—	남녀공학 구분	<제1-3차 조사> 1 남자학교 2 여자학교 3 남녀공학 <제4-6차 조사> 1 남자학교 2 여자학교 3 남녀공학이고 남자가자반이 따로 있다 4 남녀공학이고 남녀 혼합이다

변인명	문항번호							변인 설명	변인값
	제1차	제2차	제3차	제4차	제5차	제6차	제7차		
INT1C10	—	Y16-10	Y17-10	Y09	Y09b	Y06b	—	성적: 과목별 주관적 평가-미술	
INT1C11	—	Y16-10	Y17-10	Y09	Y09b	Y06b	—	성적: 과목별 주관적 평가-체육	
INT1C12	—	—	—	Y09	Y09b	Y06b	—	성적: 과목별 주관적 평가-제2외국어(일본어/프랑스어)	
INT1E	—	—	—	—	Y10	Y07	—	성적: 전체 성적 주관적 평가	전체 성적 주관적 평가 1 매우 우수 2 우수 3 약간 우수 4 보통수준 5 약간 미흡 6 미흡 7 매우 미흡
INT1D	—	Y16-2	Y17-2	Y10	Y11	Y08	—	성적: 전체 성적 만족도	전체 성적 만족도 1 매우 만족한다 2 만족하는 편이다 3 만족하지 않는 편이다 4 전혀 만족하지 않는다

- 5주기 데이터를 이용하였으므로 데이터 테이블에서 COEDUw5, INT1Ew5 속성을 찾는다.

3 데이터 가공하기

코답에서 데이터를 처리하는 방법은 다음과 같다.



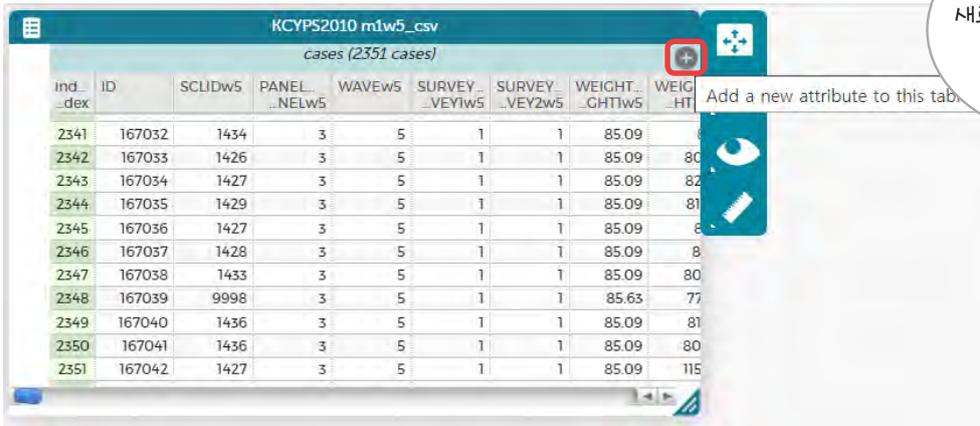
도움 자료

데이터를 가공하는 이유

원하는 정보를 습득하려고 할 때, 주어진 데이터 셋의 기본 속성만으로는 어려운 경우가 있다. 예를 들어, 새로 나온 제품에 대한 연령대별 만족도 평균 점수를 비교하고 싶다면 데이터를 가공하여 평균을 구해야 한다.

‘한국 아동·청소년패널조사 2010’ 5주기 데이터를 이용하여 데이터를 가공하는 과정을 살펴보자. 먼저 남녀공학 여부에 따라 학생이 자신의 성적에 대해 어떤 평가를 하는지 살펴보면 현재 데이터 셋에 남녀공학 여부, 학생 개인별 성적에 대한 주관적 평가를 한 점수만 속성으로 제시되어 있다는 것을 알 수 있다. 이때 남녀공학에 따른 평균 점수를 가공해야 한다.

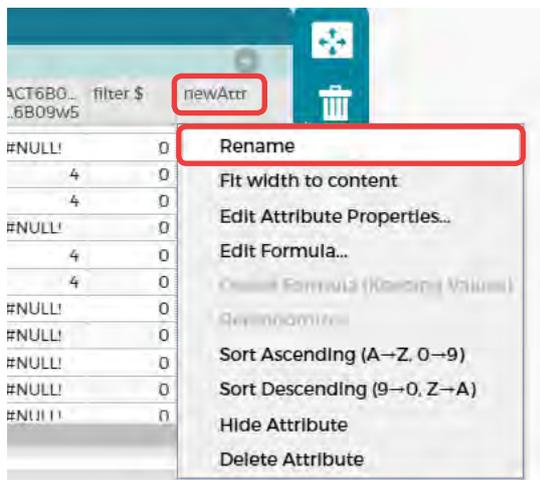
- 1 평균값을 저장하기 위한 새로운 속성 INT1Ew5_m을 만든다.
- ⊕를 눌러 속성을 하나 추가한다.



데이터 개수가 적어 새로운 속성을 생성하기까지 시간이 걸리므로 잠시 기다려야 합니다.

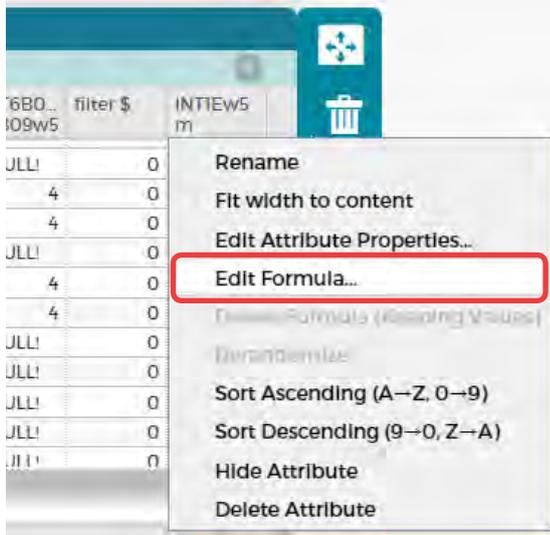


- 테이블의 제일 오른쪽에 새로운 속성 ‘newAttr’이 생성된다.
- 속성명을 클릭하고 ‘Rename’을 선택하여 ‘INT1Ew5_m’이라는 이름으로 변경한다.



② 전체 성적 주관적 평가(INT1Ew5) 속성의 평균을 구한다.

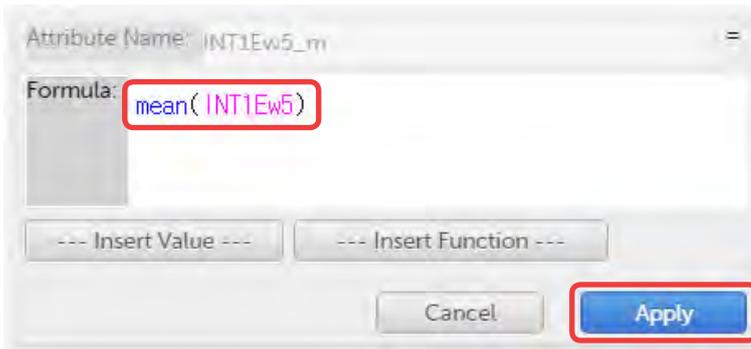
- 전체 성적 주관적 평가(INT1Ew5) 속성명을 클릭하여 'Edit Formula..'를 선택한다.



이 과정은 나중에 남녀공학 구분(COEDUw5) 속성을 기준으로 각 다른 평균으로 나누기 위한 전처리 단계이다. 속성에 평균 함수(mean)를 이용하여 전체 성적 주관적 평가(INT1Dw5)의 평균을 구합니다.



- 속성 Formula에 mean(INT1Ew5)라고 입력한 후 적용(Apply) 버튼을 클릭한다.



일반적으로 평균은 average를 많이 쓰지만 통계 분야에서는 mean이라고 쓴다.

- INT1Ew5_m 속성에 4.05값이 전부 입력되었다면 잘 수행한 것이다.

BO...	ACT6B0...	ACT6B0...	ACT6B0...	ACT6B0...	ACT6B0...	ACT6B0...	filter \$	INT1Ew5
3w5	...6B04w5	...6B05w5	...6B06w5	...6B07w5	...6B08w5	...6B09w5		m
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
4	4	4	4	4	4	4	0	4.05
4	1	1	1	1	1	4	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
4	4	4	4	2	4	4	0	4.05
4	2	4	4	2	2	4	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05
.L!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!	0	4.05

남녀공학 구분(COEDUw5)에 따라 평균값을 정리하기 위한 과정이므로 모두 같은 값 4.05가 포함되어도 상관없습니다.

③ 남녀공학 구분(COEDUw5) 속성을 기준으로 그룹화한다.

- 남녀공학 구분(COEDUw5) 속성으로 그룹화하려면 남녀공학 구분(COEDUw5) 속성을 테이블의 제일 앞(Index 속성 왼쪽)으로 끌어온다.

Ind... _dex	ID	SCLIDw5	PANEL... _NELw5	WAVEw5
2339	167030	1428	3	5
2340	167031	1429	3	5
2341	167032	1434	3	5
2342	167033	1426	3	5
2343	167034	1427	3	5
2344	167035	1429	3	5
2345	167036	1427	3	5
2346	167037	1428	3	5
2347	167038	1433	3	5
2348	167039	9998	3	5
2349	167040	1436	3	5
2350	167041	1436	3	5
2351	167042	1477	3	5

OO... OLw5	GRADE... ADEw5	COEDU... EDUw5	SCLTY... _TYPw5
3	2	2	
3	2	4	
3	2	3	
3	2	2	
3	2	4	
3	2	2	
3	2	3	
-8	#NULL!	#NULL!	
3	2	4	
3	2	4	
3	2	2	

남녀공학 구분 (COEDUw5) 속성을 Index 속성 왼쪽으로 끌어오면 Index 속성 왼쪽이 흰색에서 노란색으로 바뀝니다.



- 남녀공학 구분(COEDUw5) 속성을 기준으로 그룹이 생성된다. □를 클릭하면 각 그룹에 해당하는 데이터의 개수가 표시된다.

Ind... _dex	COEDU... EDUw5	ind... _dex	ID	SCLIDw5
1	4	1	14201	1503
2	#NULL!	2	14302	1547
3	1	3	14308	2801
4	2	4	14309	1554
5	3	5	14316	1558
		6	14324	1533
		7	14329	1534
		8	14429	1571
		9	14432	1565
		10	14523	1533
		11	14524	1534
		12	14535	1537
		13	14618	1539

Ind... _dex	COEDU... EDUw5	ind... _dex	ID	SCLIDw5
1	4			674 cases
2	#NULL!			299 cases
3	1			464 cases
4	2	1	14210	1507
		2	14212	1508
		3	14214	1509
		4	14215	1510

- 그룹별로 전체 성적 주관적 평가 점수의 평균을 보려면, 전체 성적 주관적 평가 평균(INT1Ew5_m) 속성을 남녀공학 구분(COEDUw5) 속성 오른쪽으로 드래그 앤 드롭한다.

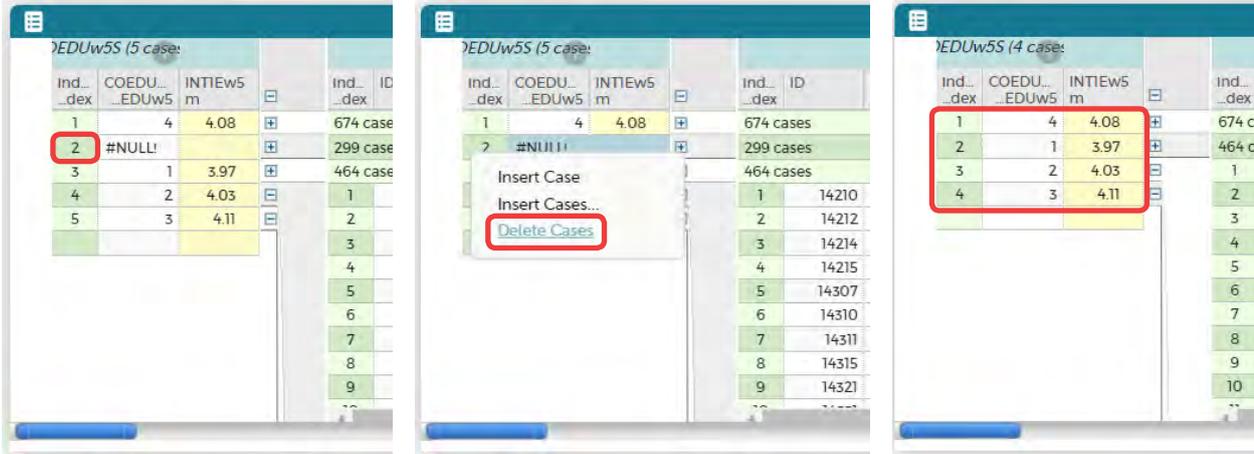
Ind... _dex	COEDU... EDUw5	INT1Ew5_m	ID	SCLIDw5	
1	4	674 cases			
2	#NULL!	299 cases			
3	1	464 cases			
4	2		1	14210	1507
			2	14212	1508
			3	14214	1509
			4	14215	1510
			5	14307	1515
			6	14310	1515
			7	14311	1515
			8	14315	1515
			9	14321	1515
			10	14331	1515

ACT6BO... _6BO9W5	filter \$	INT1Ew5_m
#NULL!	0	4.05
4	0	4.05
#NULL!	0	4.05
4	0	4.05
#NULL!	0	4.05

III 데이터

④ 빈 값 #NULL!을 삭제한다.

- 그룹화한 Index에서 '#NULL!'이 있는 번호를 클릭하여, 'Delete Cases'를 선택하면 해당 번호가 삭제된다.



4 데이터 시각화하기

남녀공학 구분(COEDUw5), 전체 성적 주관적 평가 평균(INT1Ew5_m) 속성을 이용하여 그래프를 출력한다.

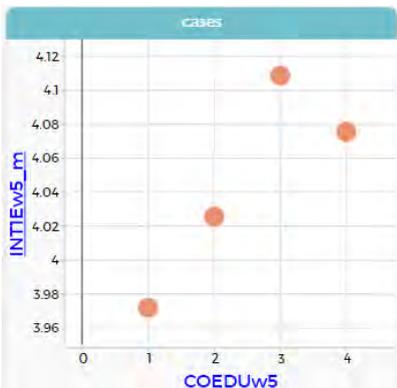
① Graph 메뉴를 선택한다.



② x축에 남녀공학 구분(COEDUw5) 속성, y축에 전체 성적 주관적 평가 평균(INT1Ew5_m) 속성을 드래그 앤 드롭한다.

5 결과 해석하기

결과 그래프를 통해 남녀공학 여부에 따른 전체 성적에 대한 주관적인 평가를 비교해 보자.



남녀공학 구분 속성의 1, 2, 3, 4는 범주형 데이터로 각 숫자는 다음을 나타낸다.
 1 = 남학교
 2 = 여학교
 3 = 남녀공학이지만, 남자반, 여자반 분리
 4 = 남녀공학이면서 남녀 합반

2. 장래학업 수준에 대한 부모님과 학생 간의 차이 분석 **심화 활동(선택)**

부모님이 생각하는 장래 교육 수준과 학생들이 생각하는 장래 교육 수준 간의 관계를 살펴보자.

1 필요한 속성을 찾아보자.

미래와 관련된 속성들은 코드북 82쪽~83쪽에 제시되어 있으며 KCYPS2010m1w6 파일을 활용합니다.



2 데이터 가공이 필요하다면, 새로 만들어 보자.

3 결측값이 있다면 이를 처리해 보자.

4 데이터를 시각화해 보자.

5 결과를 해석해 보자.

4 정리하기

1 활동을 통해, 알게 된 내용을 정리해 보자.

• 데이터를 수집하는 방법은 _____

 _____ 있다.

• 데이터 가공 목적과 데이터 가공 방법

항목	내용
데이터 가공 목적	
데이터 가공 방법	

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
데이터 가공 필요성	데이터를 가공해야 하는 필요성을 이해하였는가?					
데이터 처리를 통한 문제해결	문제해결에 필요한 데이터를 수집 및 가공하고 시각화하여 문제의 해를 얻었는가?					
참여도	활동에 적극적으로 참여하였는가?					

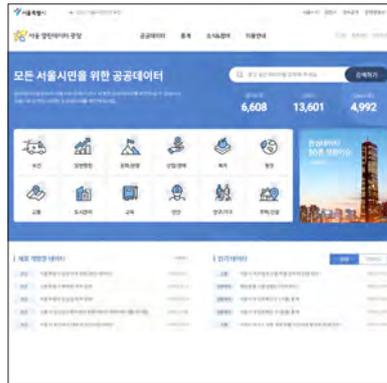
문제 해결에 필요한 데이터는 어디서 수집할 수 있는지 알아보자.

1 공공 데이터

정부나 지방자치단체, 공공기관에서 운영하는 데이터 사이트를 이용한다. 공공데이터포털이나 서울, 경기 등에서 운영하는 데이터 사이트에 여러 분야의 데이터를 무료로 제공한다. 그 외에도 외교부, 한국 아동·청소년 데이터 아카이브 사이트 등을 통해 데이터 수집이 가능하다.



▲ 공공데이터포털
<https://www.data.go.kr/>



▲ 서울 열린데이터 광장
<https://data.seoul.go.kr/>



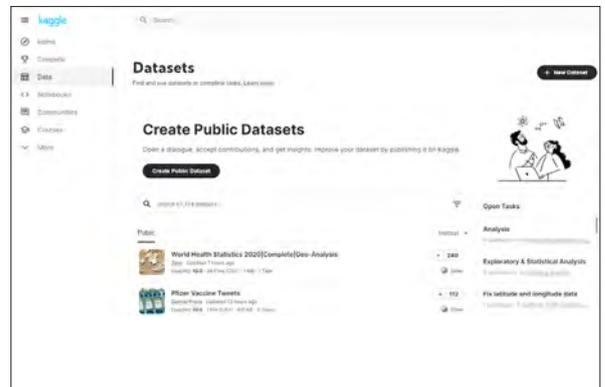
▲ 경기데이터드림
<https://data.gg.go.kr/>

2 민간 데이터

공공기관이 아닌 민간에서 운영하는 데이터 사이트를 이용한다. 경기데이터드림 사이트에는 민간 데이터 카탈로그를 별도로 제공하며 해당 데이터 사이트로 이동하여 비용을 지불하여 사용할 수 있다. 캐글(kaggle)은 데이터를 공유하고 이를 이용하는 사람들이 데이터로부터 통찰을 제공할 수 있도록 무료 데이터 공유 사이트를 제공하고 있다.



▲ 경기데이터드림의 민간데이터 카탈로그
<https://data.gg.go.kr/portal/data/catalog/searchPrivateCatalogPage.do>



▲ 캐글(kaggle)
<https://www.kaggle.com/datasets>



데이터 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

1 배운 내용을 마인드맵으로 표현하고 친구들과 비교해 보자.

마인드맵으로 표현

2 내가 만약 '데이터 분석가'가 된다면 지금 무엇을 배우고 어떤 마음가짐을 가져야 할지 써 보자.

배워야 할 부분

마음가짐

3 우리 학교에 학생들에 대한 다양한 데이터를 수집할 수 있다면 어떤 분석을 해 보고 싶은지 작성해 보고 친구들과 공유해 보자.

1 어떤 분석을 하면 좋을지 그 이유와 함께 작성해 보자.

2 학생들에 대한 어떤 데이터가 필요할지 생각해 보자.

3 학생들로부터 해당 데이터를 수집할 때, 개인 정보 보호, 사생활 보호를 위해 어떤 조치를 취해야 할지 생각해 보자.

IV 기계학습

1 단계 핵심 개념 알기

- 기계학습
- 인공지능 프로젝트 사이클



2 단계 창의융합적 사고로 해결하기

- 💡 1 기계학습으로 무엇을 할 수 있을까?
- 💡 2 인공지능으로 해결할 수 있는 문제를 찾아볼까?
- 💡 3 기계학습 모델이 학습할 데이터는 어떻게 모을까?
- 💡 4 나의 모델은 얼마나 똑똑해졌을까?

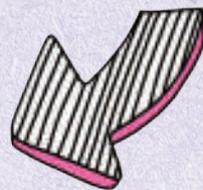
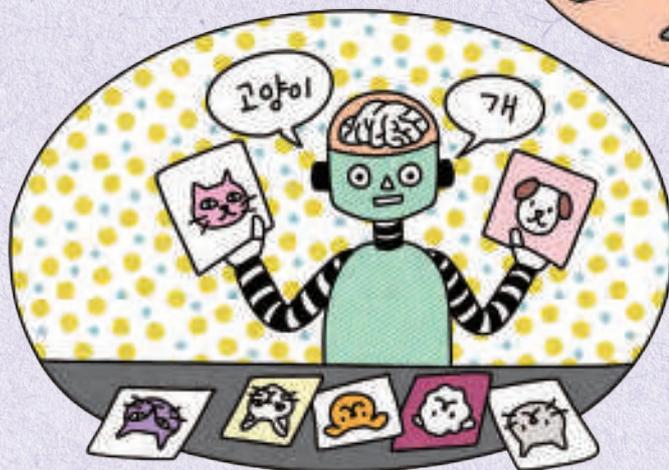
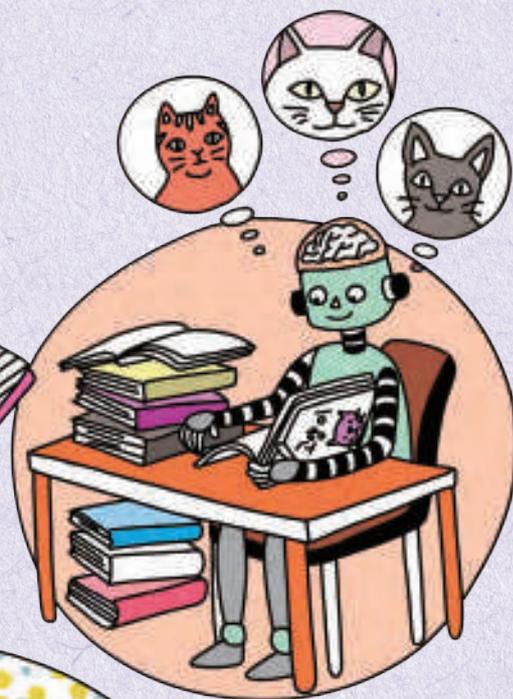
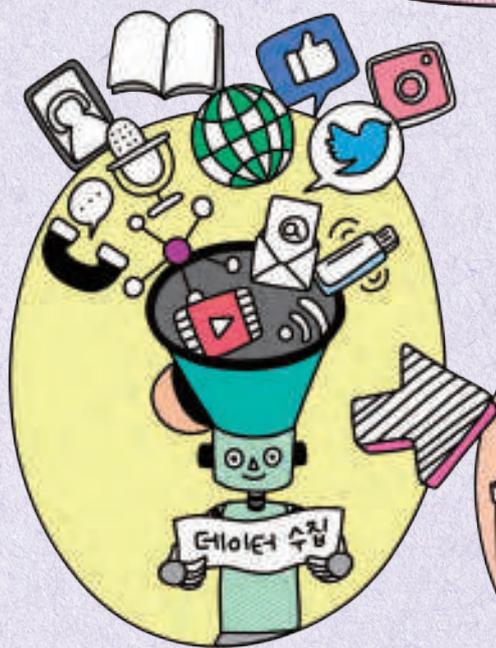


3 단계 공감 나누기

- 나만의 인공지능 프로젝트 만들기

③ 기계학습 단원에서는 무엇을 배울까요?

- 문제 해결에 필요한 데이터를 준비하여 훈련과 테스트 과정을 통해 성능을 평가하고 분류 모델을 구현하여 문제를 해결합니다.





1단계 핵심 개념 알기

기계학습과 인공지능 프로젝트 사이클

✓ 기계학습의 형태를 구분하고 인공지능 프로젝트 개발 과정을 설명할 수 있다.

기계학습은 인공지능의 한 분야로서 사람의 직접적인 명령이 없어도 데이터를 통한 학습 경험을 바탕으로 성능을 발전시킨다. 기계가 데이터를 통해 학습한다는 의미에서 ‘기계학습(Machine Learning)’이라고 한다. 본 단원에서는 학습 방법에 따른 기계학습의 종류를 알아보고, 문제 정의부터 배포와 활용까지의 체계적인 인공지능 프로젝트 사이클을 적용하여 문제를 해결한다.

1 기계학습

Q 개념 탐색 기계학습은 무엇일까?

기계학습은 학습 방법에 따라 크게 두 가지 방법으로 구분한다.

1 지도 학습

데이터에 대한 명시적인 정답 즉, 레이블(Label)이 주어진 상태에서 컴퓨터가 학습하는 방식이다. 레이블을 붙여놓은 데이터를 사람이 컴퓨터에 입력하면서 “이것은 무엇이야.”라고 정의하면 컴퓨터 입장에서는 컴퓨터가 지도(Supervised)를 받았기 때문에 이를 지도 학습(Supervised Learning)이라고 말한다.

2 비지도 학습

지도 학습과 반대로 “이것은 무엇이야.”라고 지도를 받지 않고 레이블(Label)이 없는 데이터를 제공받아 새로운 데이터에 대한 결과를 예측하는 것을 비지도 학습(Unsupervised Learning)이라고 한다.

2 인공지능 프로젝트 사이클

Q 개념 탐색 인공지능 기술을 이용한 프로젝트는 어떤 과정을 거쳐 개발할까?

기계학습 프로젝트 사이클에서 각 단계는 경우에 따라 생략될 수도 있으며 필요에 따라 반복될 수도 있습니다.



“기계학습 모델은 문제를 정의하고, 데이터를 준비하여
훈련과 테스트 과정을 통하여 구현된다.”

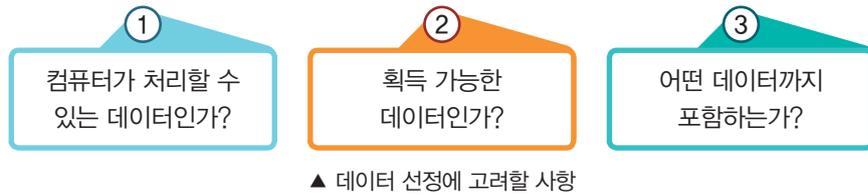
인공지능으로 문제를 해결하기 위해서는 특정 단계를 거쳐 수행된다. 인공지능 프로젝트 사이클은 문제를 정의하고 해결할 문제의 범위를 설정하는 것부터 시작하여 모델을 평가하고 배포 및 활용하는 단계로 진행된다.

1 문제 정의

인공지능 프로젝트를 시작하는 가장 중요한 단계로서 우리가 인공지능 기술을 이용하여 해결하고자 하는 프로젝트의 목표를 정의한다.

2 데이터 수집 및 전처리

인공지능 모델 훈련에 필요한 데이터를 수집하고 전처리하는 단계이다.



3 데이터 탐색

데이터의 특성을 파악할 수 있도록 그래프나 차트 등의 친숙한 형태로 시각화하고 결측치나 이상치 등을 제거하는 정제 작업을 포함한다.

4 모델 학습

다양한 인공지능 모델 중 해결하고자 하는 문제에 적합한 모델을 선정하여 모델을 만들고 학습시키는 단계이다.

5 모델 평가 및 배포

문제 해결을 위해 구현한 다양한 모델을 평가하고 최상의 모델을 선택하고, 다수의 사람들이 활용할 수 있도록 공유한다. 이 단계는 우리가 인공지능 기술을 배우는 궁극적인 이유라고 할 수 있다.

활동 미리 보기

핵심 개념과 관련된 활동은 어떤 내용으로 전개될까요?

활동 1

기계학습으로 무엇을 할 수 있을까?

- 지도 학습
- 비지도 학습
- 분류, 군집

활동 2

인공지능으로 해결할 수 있는 문제를 찾아볼까?

- 문제의 설정
- 4Ws

활동 3

기계학습 모델이 학습할 데이터를 어떻게 모을까?

- 데이터 수집
- 데이터 전처리

활동 4

나의 모델은 얼마나 똑똑해졌을까?

- 모델 학습
- 모델 성능 평가



2단계 창의융합적 사고로 해결하기

활동 1 기계학습으로 무엇을 할 수 있을까?

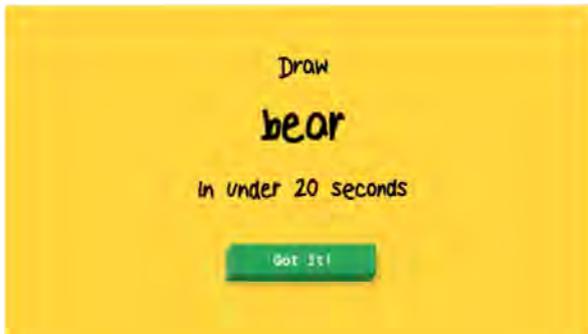
- ▶ 기계학습을 이용한 게임에서 사용한 기법이 분류, 군집 중 어디에 속하는지 분류할 수 있다.

1 문제 이해하기

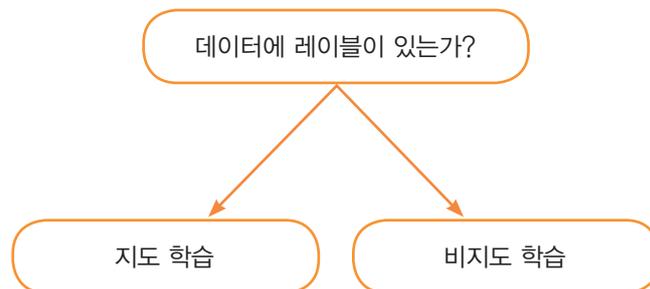
구글의 Experiments with Google 게임에서 인공지능 기술을 찾아보자.

구글의 Experiments with Google 사이트의 다양하고 재미있는 게임들은 대부분 기계학습 기술을 이용하고 있다.

그중에 Quick, Draw!라는 게임을 해 보자. 이 게임은 컴퓨터에서 제시하는 대상을 사람이 그림으로 그리면 그 그림이 무엇인지를 인공지능 시스템이 맞추는 게임이다.



이 게임에서 사용한 기계학습 모델은 지도 학습과 비지도 학습 중 어디에 속할까? 앞서 배운 대로 지도 학습인지, 비지도 학습인지 파악하려면 데이터에 레이블이 있는지 확인해야 한다.



Quick Draw의 그림에서는 컴퓨터가 그리라고 제시한 대상 즉, 위 그림에서는 곰(bear)이 레이블에 해당한다. 이 게임에서는 그림의 레이블이 존재하므로 지도 학습 방법을 이용했다고 볼 수 있다.

구글의 Experiments with Google 사이트의 또 다른 게임을 살펴보자.

무엇을 알아야 할까?

배경 지식

분류, 군집

1. 분류

지도 학습에서는 아래 그림처럼 이름, 즉 레이블(Label)을 붙여놓은 데이터를 이용한다.

							
불 테리어	시바견	소말리	재패니즈 밥테일	닥스훈트	러시안 블루	불도그	페르시안

▲ 레이블

지도 학습에서 사용되는 대표적인 모델에는 분류(Classification)가 있다. 분류는 일반적으로 결괏값, 즉 레이블이(Label)이 몇 가지 종류로 고정된 범주형 문제에 사용한다. 예를 들어 개, 고양이 또는 사과, 배, 귤 그리고 객관식 시험의 답처럼 1, 2, 3, 4, 5로 레이블이 주어진 데이터를 학습할 때 주로 사용한다.

위 그림과 같이 강아지의 종류를 레이블로 달고 있는 강아지 사진을 학습한 뒤, 사진을 보여주면 해당하는 강아지의 종류, 즉 레이블을 찾아주는 시스템이 분류 모델을 사용하는 인공지능 시스템이다.



2. 군집

레이블(Label)이 없는 데이터를 사용하는 비지도 학습은 군집(Clustering) 모델이 있다.

군집은 수많은 자료들을 특성에 따라 몇 가지 그룹으로 만드는 것이다. 비지도 학습의 데이터는 레이블(Label)이 없기 때문에, 무엇을 근거로 묶어서 분류하느냐가 최대 관건이 된다. 즉, 군집은 전체를 어떻게 묶어 주어야 할지 그 방법을 제시하는 것이다. 위 예제에서 보면, 귀가 뽀족한 것 - 뽀족하지 않은 것으로 묶을 수도 있고, 한쪽 눈에 반점이 있는 것 - 반점이 없는 것으로 묶을 수도 있다.

2 문제 살펴보기

구글 Experiments with Google의 다양한 게임을 체험하고 그 게임이 어떤 모델을 사용했는지 토론해 보자.

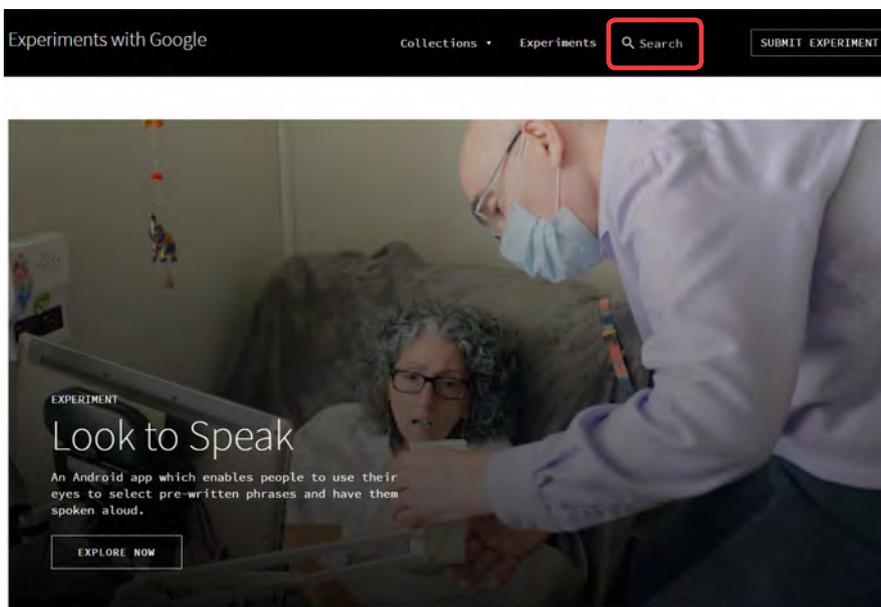
1 Experiment with Google 사이트 접속하기

주소 창에 URL을 입력하여 접속한다. 이때, 크롬() 브라우저에서 접속해야 오류 없이 사용할 수 있다.

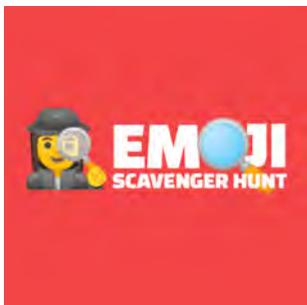
<https://experiments.withgoogle.com/>

2 게임 검색 후 체험하기

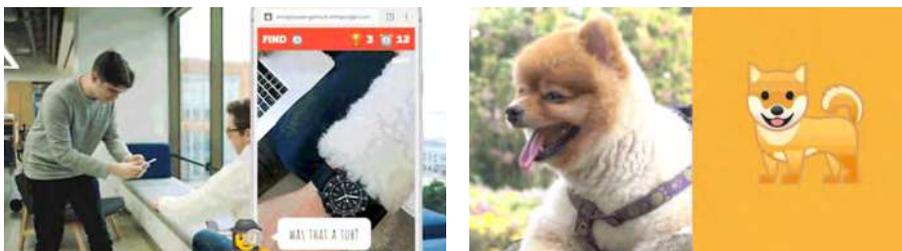
검색창을 이용하여 게임을 검색하고 체험한다.



1 Emoji Scavenger Hunt(이모지 스캐벤저 헌트)



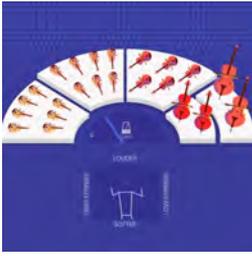
카메라를 이용하는 게임으로 게임에서 말하는 오브젝트 (책, 휴대전화 등)를 가져와 카메라 앞에 보여 주면 그 물체를 인식하는 게임이다. 시간 제한이 있으므로 빠르게 물건을 찾아서 카메라에 보여 줘야 한다.



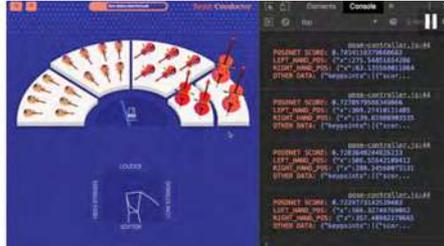
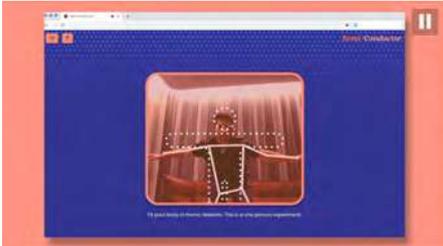
EMOJI는 PC와 모바일에서 모두 작동합니다.



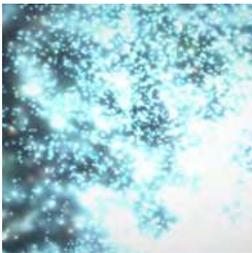
② Semi-Conductor(세미 컨덕터)



세미 컨덕터는 브라우저를 통해 직접 오케스트라를 지휘하는 게임이다. 카메라 앞에서 팔을 움직여 브라우저를 통해 자신만의 오케스트라를 지휘할 수 있는 게임으로 음악의 템포, 볼륨, 악기를 변경할 수 있다. 이 프로그램은 PoseNet이라는 기계학습 라이브러리를 이용하였다.



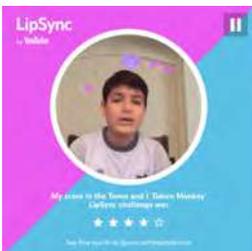
③ Kluster(클러스터)



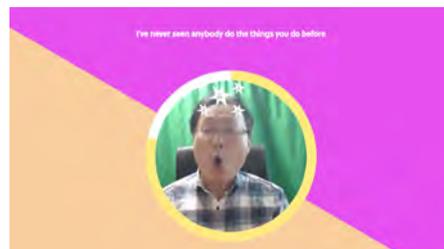
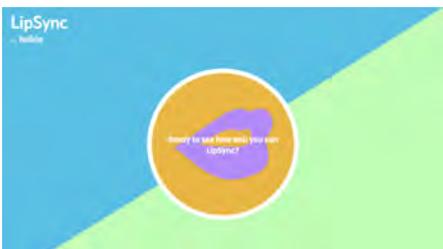
사용자가 가지고 있는 그림이나 게임에서 제공하는 샘플 사진을 선택하면 사진에서 사용한 색상을 이용해서 우주 공간의 은하수 같은 형태의 그림을 만들어 준다. 아래와 같이 게임의 왼쪽 메뉴 중 [NEW IMAGE → SELECT FROM DEVICE]를 이용하여 내 컴퓨터에 있는 사진을 선택하면 사진 속 색상을 이용하여 은하수 사진을 만들어 준다.



④ LipSync(립싱크)



우리가 예전에 즐기던 여러 가지 박자 관련 게임이나 댄스 게임과 비슷한 게임이다. 게임에서 제공하는 노래를 립싱크하는 게임으로 컴퓨터에 장착된 카메라 앞에서 노래를 하면 립싱크를 인식한다. 립싱크를 잘하면 오른쪽 사진처럼 별도로 장식해 주고 왕관도 씌워 준다.

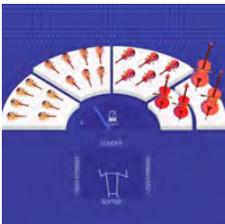


3 게임 속 기계학습 기술 분류하기

- ① 조별 토론을 통하여 어느 게임이 분류 모델을 사용했고 어느 게임이 군집 모델을 사용했는지 찾아보고 표에 게임의 이름을 적어보자.

분류	군집

- ② 각 게임을 분류한 이유와 특징을 조별로 적어보고 다른 조와 비교해 보자.

게임	사용된 모델	특징
 <p>Emoji Scavenger Hunt</p>		
 <p>Semi-Conductor</p>		
 <p>Kluster</p>		
 <p>LipSync</p>		

3 토론하기

1 다음의 사례에서 사용한 기계학습 기술은 분류, 군집 중 어디에 속하는지 토론해 보자.

美 뉴욕주, 학교에서 얼굴 인식 기술 사용금지



미국 뉴욕주가 학교에서 얼굴 인식 사용을 금지했다.

그동안 미국 여러 주에서는 정부 및 공공 기관에서 얼굴 인식 기술을 사용하지 못하도록 했다. 하지만 학교에까지 얼굴 인식 사용 금지 명령을 내린 것은 뉴욕주가 처음이라고 미국의 IT 전문 뉴스 미디어 엔가젯이 23일(현지시간) 보도했다.

뉴욕주는 얼굴 인식 기술 사용 금지 시한도 함께 규정하였는데, 2022년 7월 1일, 또는 사생활 침해나 보안 문제 등에 대해 문제가 없다는 공식 확인이 나올 때까지 금지한다고 밝혔다.

(후략)

[출처: <https://zdnet.co.kr/view/?no=20201224082307>]

2 현재 사용하는 인공지능 기술 중 분류에 해당하는 기술의 예를 찾아보자.

●
●
●
●
●
●
●
●

3 분류 모델을 이용하여 미래 적용할 인공지능 시스템을 제시해 보자.

●
●
●
●
●
●
●
●

4 정리하기

1 활동을 통해 알게 된 내용을 정리해 보자.

2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
지도 학습과 비지도 학습	지도 학습과 비지도 학습의 차이를 설명할 수 있는가?					
분류	분류 모델의 특징을 설명할 수 있는가?					
참여도	활동에 적극적으로 참여하였는가?					

AI 발전에 앞장서고 있는 IBM은 최근 'Project Debater(이하, Debater)'를 개발하여 토론 영역까지 AI를 확대하고 있다. Debater는 인간과 복잡한 주제를 토론할 수 있는 AI로, IBM은 인간보다 설득력 있는 주장과 더불어 정보에 입각한 결정을 내리는 데 도움을 주기 위해 Debater를 개발했다.



실제로, 2019년 2월 Debater는 2016년 국제 토론 대회 결승 진출자인 하리시 나타라잔(Harish Natarajan)과 '유치원에 보조금을 지급해야 하는가?'라는 주제로 20분간 열띤 토론을 펼쳤다. 이 주제에 대해 Debater는 찬성을 선택해, 정부가 유치원에 보조금을 지급하게 되면 소외 계층의 교육 기회가 확대되고, 교육의 형평성으로 인해 교육 수준이 강화되어 궁극적으로 지역 사회에 도움이 될 것이라고 논증했다. 반면 나타라잔은 보조금 지급이 이루어져도 모든 어린이들이 유치원에 다닐 수 없다는 사실을 지적했다.

과연 누구의 논리가 승리했을까?

그뿐만 아니라, Debater는 인공지능을 유익하게 사용하기 위해 UN의 글로벌 플랫폼인 'AI for Good'의 2019년 글로벌 정상회담에 대표 AI로 초대되어 국제 사회의 높은 관심을 받았다.

회담을 통해, 국제 사회는 Debater가 사회 문제 해결 및 지속가능개발목표(SDGs) 달성을 위해 효과적인 정책과 방향성을 결정해야 할 때 도움을 줄 것으로 전망했다.

17개의 목표와 169개의 세부 목표로 구성된 SDGs를 한정적인 자원 아래 2030년까지 성공적으로 달성하기 위해서는 바른 결정에 따른 전략적 접근이 요구된다. 이에 있어, 세상에 존재하는 정보 데이터 분석으로 옳고 그름을 따져 논리적인 판단을 제시할 수 있는 Debater가 과학적이고 타당한 결정에 도움을 줄 수 있다.

이에 발맞추어, IBM 또한 Debater를 의사 및 정책 결정에 활용할 수 있는 시스템으로 강화시켜, 기업과 정부에 제공하고자 계획 중이다. 따라서 더 나은 세상으로 나아가기 위한 인류의 결정과 판단에 Debater와 같은 AI가 앞으로 큰 도움을 줄 수 있을 것이라고 예상된다.

[출처: <https://blog.naver.com/innosociallab/221847033823> 발췌]

활동 2

인공지능으로 해결할 수 있는 문제를 찾아볼까?

▶ 주변에 있는 문제 상황을 이해하고 인공지능을 이용하여 문제를 해결할 수 있다.

1 문제 이해하기

다음 글을 읽고, 세상의 다양한 문제 중 나는 어떤 문제에 관심을 가지고 있는지, 그 문제는 인공지능 기술을 사용할 수 있는 문제인지 알아보자.

안과 진료를 제때에 받지 못하거나 자신의 시력에 맞는 안경을 구하지 못해 시력을 잃는 사람들이 전 세계적으로 약 3억만 명 이상이라고 한다. 이러한 사례는 대부분 경제적, 정치적 상황 등으로 의료 인프라가 제대로 갖추어지지 않은 개발도상국에서 발생한다고 한다.

여러 단체에서 이런 어려움을 가지고 있는 국가에 의료 지원을 보내고 『국경없는 의사회』 등에서도 적극적인 의료 봉사활동을 하고 있지만 상황을 개선하기에는 많이 부족하다. 특히, 안과 검진의 경우 상당수 의료 봉사나 국제적 지원이 부족한 실정이다.



▲스마트폰 앱을 이용한 눈 건강 검진(출처: peekvision)

이러한 상황을 고려하여 한 단체에서는 스마트폰 앱을 이용하여 눈 건강 상태를 검진할 수 있는 인공지능 기반 프로그램을 제시하였다.

안과 검진을 제때 받지 못하는 빈곤 국가의 이웃들을 위해 개발된 이 애플리케이션에서는 스마트폰의 카메라를 통해 검진이 이루어진다. 카메라에서 나오는 플래시 불빛의 자극에 눈이 어떻게 반응하는지를 분석하고 다양한 각도에서의 사진과 영상 자료를 저장한다. 이렇게 저장된 검진 자료들은 색맹 테스트부터 백내장 등과 같은 심각한 안구 질환까지 이미지 인식 등의 인공지능 기술을 이용하여 진단한다. 의사는 이를 통해 다시 한번 환자의 안구 건강 상태를 진단하고 처방을 내려 환자의 눈 건강을 지속적으로 관리한다.



▲스마트폰 앱을 이용한 눈 건강 검진(출처: peekvision)

문제는 누구와 관계된 것이며, 왜 해결할 필요가 있는지 생각해 보자.

🎯 무엇을 알아야 할까?

🔍 배경 지식

문제의 설정

인공지능 기술을 이용하여 해결할 수 있는 문제들은 매우 많다. 이미 상용화 단계에 이른 자율주행자동차라든지, 벌써 수십 년 전부터 사용해 온 주차장 번호판 자동 인식 시스템 등이 대표적인 사례이다.

구글은 Euphonia라는 프로젝트를 통하여 청각 등의 문제로 이례적인 발음을 하는 사람들의 발음을 인식해 주는 시스템을 개발하였다. 또한 PeekVision이라는 기업은 안과 진료를 받기 어려운 사람들이 시력을 잃는 것을 막기 위해 휴대전화의 카메라를 이용하여 눈의 상태를 진단하는 데 도움을 주는 인공지능 프로그램을 개발하여 보급하고 있다.

의사의 치료에 도움을 주는 눈 건강 검진 앱은 국제 눈 건강센터에서 개발한 무료 앱으로 의료 사각지대에 있는 많은 사람들에게 도움을 줄 수 있을 것으로 생각된다.

이렇게 인공지능은 어떤 문제를 위해 어떻게 활용하는지가 중요하다.



Project Euphonia

Project Euphonia is a Google Research initiative focused on helping people with atypical speech be better understood. The approach is centered on analyzing speech recordings to better train speech recognition models.

Helping people be better understood

For millions of people around the world whose speech is difficult for others to understand, face-to-face communication can be very challenging. Using voice-activated technologies can be frustrating, too. While tools like Google Home or the Google Assistant can help people call someone, adjust lighting, or play a favorite song, they may not work as well for those with impaired speech.



▲구글 Project Euphonia

2 현재 인공지능 기술을 활용하여 해결할 수 있는 문제와 그렇지 않은 문제 알아보기

인공지능으로 해결할 수 있는 문제의 사례를 살펴보자.

1 인공지능 기술로 해결 가능한 문제 살펴보기

인공지능으로 해결할 수 있는 문제는 다양한 분야에 있다.

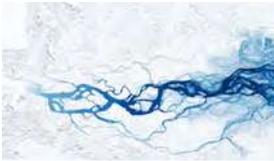
UN SDGs(Sustainable Development Goals, 지속 가능 발전 목표)의 17개 항목에 대하여 관심을 가지고 이해하면 우리가 사는 세상에 도움을 줄 수 있는 인공지능 프로젝트를 좀 더 쉽게 찾을 수 있다.



▲ 유엔의 지속 가능 발전 목표(SDGs)

2 현재 인공지능 기술을 활용하여 해결할 수 있는 문제와 그렇지 않은 문제를 구분하기

다음 사례 중 현재 인공지능을 활용하여 해결할 수 있는 문제는 O표, 해결 불가능한 문제 또는 그럴 필요가 없는 문제는 X표로 구분하고 그 이유를 써 보자.

문제	인공지능으로 해결 가능: O 해결 불가능 또는 필요 없음: X
 <ul style="list-style-type: none"> • 홍수 예측 시스템 홍수가 언제 어디서 발생할 수 있을지 예측 및 경보를 제공하는 시스템 	
 <ul style="list-style-type: none"> • Lotto 번호 예측 인공지능 지금까지의 로또 당첨 번호를 데이터로 이용하여 다음 주 로또 당첨 번호를 예측하는 시스템 	
 <ul style="list-style-type: none"> • 기후 변화의 결과 시각화 극심한 기상 이변 후 개인 주택의 스트리트 뷰 이미지를 생성하여 기후 변화를 체험하게 하는 시스템 	
 <ul style="list-style-type: none"> • AI 수어 번역기 청각 장애인이 표현하는 수어를 텍스트로 번역한다. 공공장소에서 청각 장애가 있는 사람들에게 더 나은 서비스를 제공하고 정보 접근성을 높일 수 있는 시스템 	
 <ul style="list-style-type: none"> • 지하철 승하차 분석 지난해 지하철 승하차 정보를 이용하여 시간 대별로 가장 많이 타고 내린 지하철역을 찾아 지하철 서비스 개선에 도움을 주는 시스템 	

3 나의 문제 설정하기

다음 만화를 읽고, SDGs 각 영역의 문제를 바탕으로 자신의 프로젝트 주제를 선정해 보자. 내가 선정한 프로젝트의 주제가 수치 데이터를 이용한다면  나경제의 프로젝트를, 이미지 데이터를 이용한다면  나우리의 프로젝트를 따라가면서 실습해 보자.



내가 선정한 프로젝트 주제

1 4Ws Template 참고하기

큰 주제를 정했으면 다음과 같은 4Ws Template를 참고하여 자신의 인공지능 프로젝트가 누구와 관련이 있는지(who), 해결할 문제가 무엇인지(what), 어떤 상황에서 발생하는 문제인지(where), 해결 이유와 목적이(why) 무엇인지를 고민하며 표를 만든다.



나경제의 프로젝트

who	누구	카드 회사들은
what	무슨 문제	가짜로 발생하는 카드 거래 데이터 발생을 막기 어려운
where	어떤 맥락/상황에서	전국에서 발생하는 수십 건의 카드 거래에서
why	해결 이유/목적	데이터에서 가짜로 의심되는 거래 데이터를 예측



나우리의 프로젝트

who	누구	손과 발을 자유롭게 움직일 수 없는 장애를 가진 사람들은
what	무슨 문제	전동 휠체어를 손 등으로 제어할 수 없어.
where	어떤 맥락/상황에서	휠체어를 타고 밖을 자유롭게 돌아다니고 싶은데
why	해결 이유/목적	얼굴의 방향만으로 움직임 조작이 되면 좋겠어.

2 나의 4Ws Template 작성하기

who	누구	
what	무슨 문제	
where	어떤 맥락/상황에서	
why	해결 이유/목적	

4 정리하기

1 인터넷을 이용하여 SDGs 17개 영역에 속하는 인공지능 프로젝트를 1개씩 찾아 보자.

• 내가 찾은 지속 가능 발전 목표 영역

2 현재 인공지능 기술로는 해결이 어렵지만, 미래에는 해결 가능한 문제는 무엇일지 생각해 보자.

3 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
문제의 설정	인공지능 프로젝트를 위한 4Ws 템플릿을 사용할 수 있는가?					
인공지능의 활용	인공지능을 우리 삶에 유용하게 활용한 사례를 찾을 수 있는가?					
참여도	활동에 적극적으로 참여하였는가?					

활동 3

기계학습 모델이 학습할 데이터는 어떻게 모을까?

- 필요한 데이터를 수집하여 다양한 기계학습 플랫폼에서 사용할 수 있다.

1 문제 이해하기

- 앞 활동의 나우리가 도전하는 프로젝트와 관련한 기사를 읽고, 데이터를 수집하는 사례를 살펴보자.

사용자는 윌리 7 키트를 통해 10가지 다른 표정을 선택하여 앞으로 이동, 방향 전환 및 정지 등의 전동 휠체어 제어를 할 수 있다

인텔의 인공지능(AI) 기술을 탑재한 HOOBOX Robotics가 신체에 부착하는 센서 없이 AI와 카메라를 이용하여 사용자에게 독립성과 위치 제어 능력을 제공하는 '윌리 7 키트(Wheelie 7 kit)'를 통해 사지마비 환자들은 단순한 얼굴 표정으로 10가지 다른 표정을 지으면 앞으로 이동, 방향 전환 및 정지 등의 전동 휠체어 제어를 할 수 있다.



▲ 다른 표정 인식 예

애나 베스키(Anna Bethke) 인텔 '사회를 위한 AI (AI for Social Good)' 총괄은 "세계 장애인의 날을 맞아, AI 기술이 장애인들의 삶 속에서 어떻게 그들에게 이동성과 제어력을 되찾게 해 주는지 생각해 보는 것은 의미 있는 일이다. HOOBOX Robotics의 윌리 7 키트는 AI를 활용해 이동이 제한된 사람들이 자연스러운 얼굴의 움직임을 통해 이동할 수 있도록 한 좋은 예시이다."라고 밝혔다.

[출처: 인공지능신문(<http://www.aitimes.kr>)]



▲ AI 기술을 탑재한 인공지능 전동 휠체어

프로젝트에 필요한 데이터를 획득할 방법을 생각해 보자.

🎯 무엇을 알아야 할까?

🔍 배경 지식

데이터 획득 및 탐색

기계학습은 대량의 데이터를 필요로 한다. 대량의 데이터를 수집하는 방법은 공공 데이터를 이용하는 방법, Kaggle과 같은 민간 데이터를 이용하는 방법 그리고 직접 수집하는 방법 등 다양하다. 수집된 데이터는 내가 만들고자 하는 인공지능 시스템에 꼭 적합한 데이터만 있는 것은 아니기 때문에 수집한 데이터 중에는 빠진 데이터나 잘못된 데이터가 포함되어 있을 수 있다.

• 일반적으로 수집된 데이터에서 발생하는 문제는 다음과 같다.

- 일부분이 빠진 데이터(missing)
- 데이터의 값이 터무니없는 경우
- 데이터가 편향적인 경우

아래 표에서 일련번호 3번의 체중 데이터는 없다. 또한, 4번의 혈압과, 5번의 허리둘레는 값이 정상적이지 않다. 데이터를 수집한 뒤에는 데이터를 확인하여 빠진 데이터나 터무니없는 이상치(Outlier)가 없는지 확인해야 한다.

일련번호	키	체중	허리둘레	혈압
1	165	64	32	132
2	174	62	30	110
3	155		34	150
4	170	64	30	260
5	167	79	14	160

이상치나 빠진 데이터는 기본적으로 그 값을 포함한 행을 삭제하는 것이 좋다. 하지만 데이터가 많지 않을 경우에는 그 속성의 평균값, 중간값 등을 넣어서 임의로 값을 배정하기도 한다.

이미지 데이터도 마찬가지로 작업이 필요하다. 예를 들어, 고릴라 데이터를 수집하는데 오랑우탄을 넣으면 당연히 안 된다. 더불어 키, 체중, 허리둘레를 이용한 혈압 예측 회귀 시스템을 만들 때, 고혈압 대상자는 한 명도 없이 정상 혈압인 사람만 수집한다든지, 사자 사진을 수집할 때 암사자만 수집하는 등 데이터가 편향되게 수집해서도 안 된다.



2 데이터 수집 플랫폼 이해하기

이번 활동에서는 Kaggle에서 나경제의 신용카드 관련 데이터를 수집하는 방법과 Teachable Machine에서 나우리의 인공지능 휠체어 프로젝트를 위한 얼굴 방향 데이터를 직접 수집하는 방법을 알아보자.

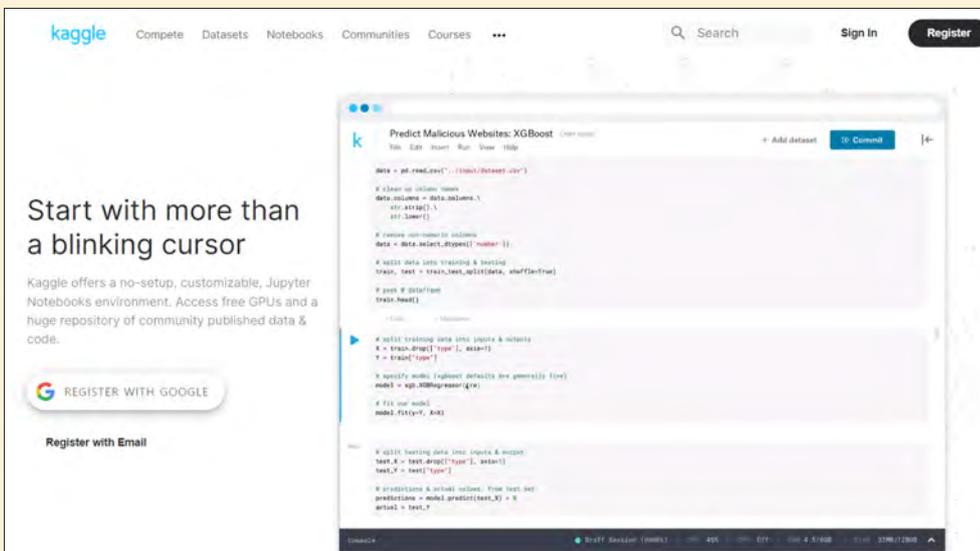


나경제의 프로젝트

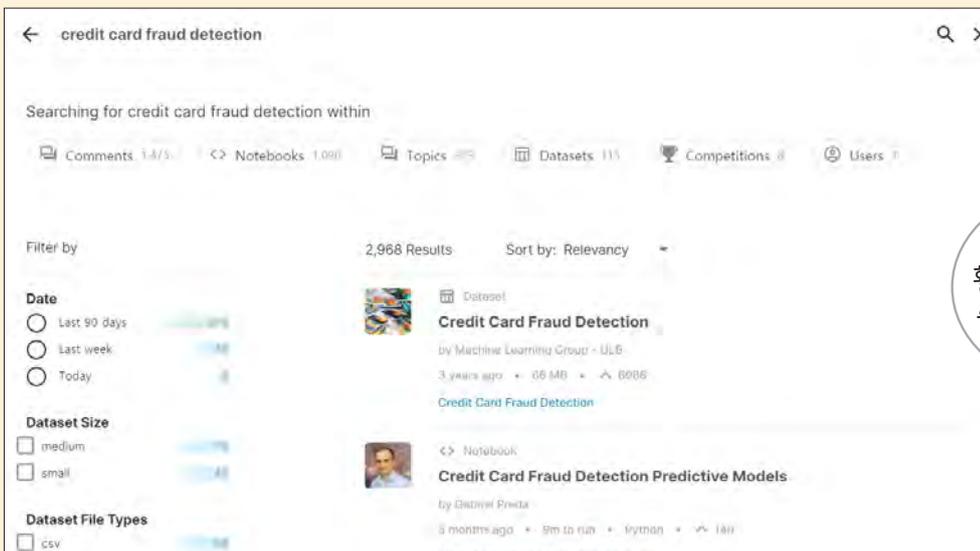
1 Kaggle에서 데이터 수집하기

- 주소창에 URL을 입력하여 접속한다.

<https://www.kaggle.com/>



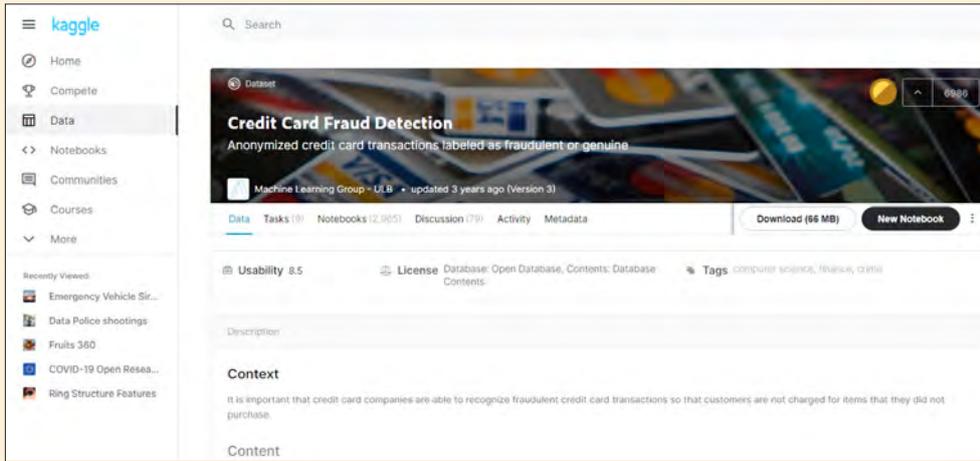
- 로그인 후 우측 상단의 'Search'창을 이용하여 'credit card fraud detection'을 검색한다.



캐글과 티처블 머신을
활용하여 직접 데이터를
수집하는 방법을 비교해
보세요.



- ③ 검색 결과가 표시된 목록 중 가장 위의 것을 선택하면 아래 그림 같이 해당 프로젝트 창으로 넘어간다.



- ④ 여기서 'Download' 버튼을 이용하면 해당 프로젝트에서 사용한 데이터를 다운로드한다.

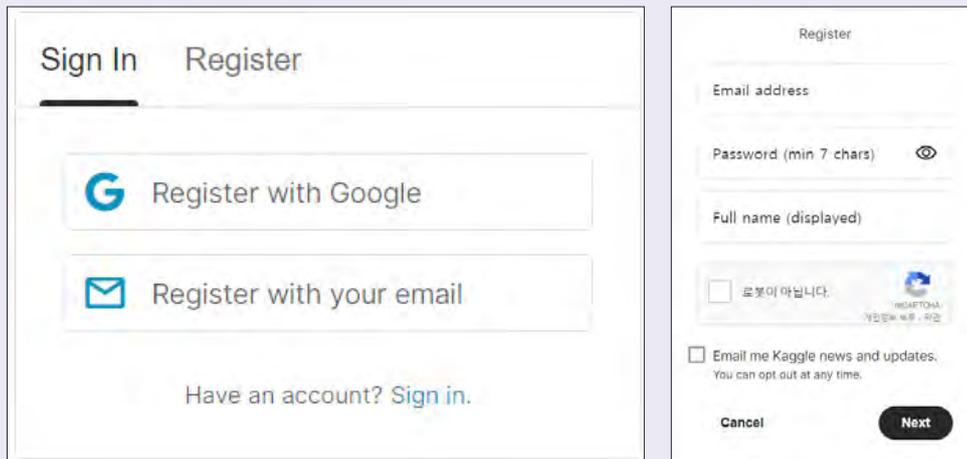
도움 자료

Kaggle(캐글)

Kaggle은 2010년에 설립된 예측 모델 및 분석 대회 플랫폼이다. 필요한 기업이나 단체에서 머신러닝으로 해결할 데이터와 과제를 등록하면 대회에 참여한 사람들이 머신러닝 모델을 개발하고 성능이 가장 좋은 모델을 만든 팀에서 수상한다. 2017년부터 구글이 인수하여 운영하고 있다.

Kaggle은 구글이 운영하는 사실 플랫폼이며 사용하려면 가입을 해야 한다.

처음 화면에서 Resister를 누르고 구글 메일 또는 개인 메일 주소만 입력하면 누구든 쉽게 가입할 수 있다.





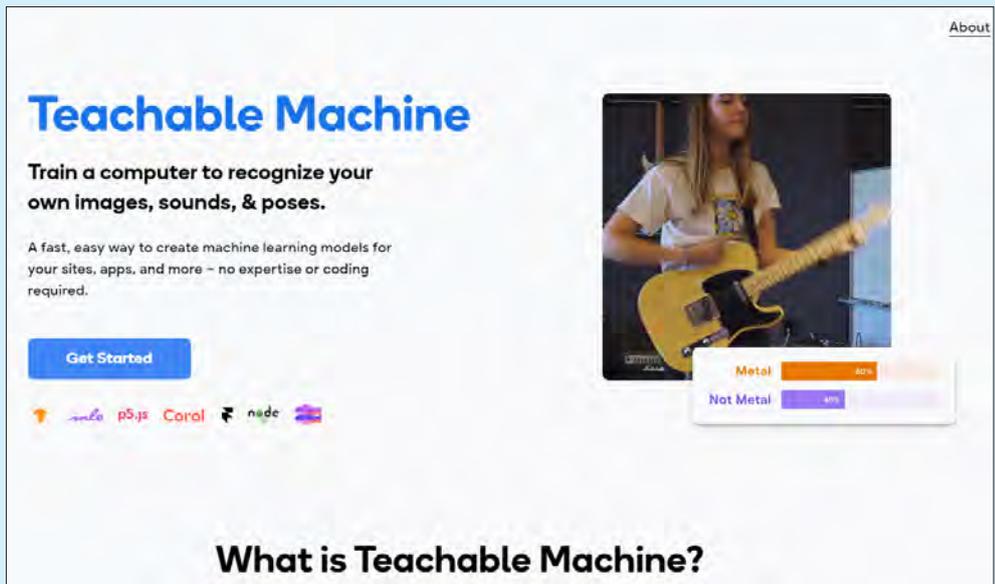
나우리의 프로젝트

2 Teachable Machine에서 데이터 직접 수집하기

- ① 주소창에 URL을 입력하여 접속한다.

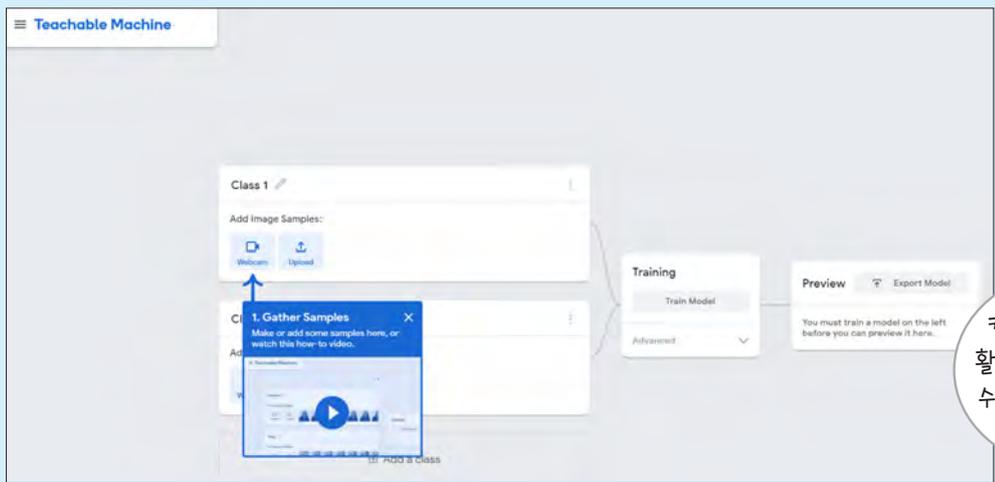
`https://teachablemachine.withgoogle.com/`

- ② **Get Started** 버튼을 누른 뒤  (Image Project)를 선택한다.



Teachable Machine은 코딩 경험이 없어도 자신만의 기계학습 모델(이미지, 사운드, 포즈)을 빠르고 쉽게 만들 수 있는 웹 기반 도구이다.

- ③  (Webcam)을 이용하여 직접 사진을 찍을 수 있고,  (Upload)를 이용하여 찍어 놓은 이미지를 업로드할 수도 있다.



캐글과 티저블 머신을 활용하여 직접 데이터를 수집하는 방법을 비교해 보세요.



3 데이터 수집과 분석하기

나경제와 나우리가 수집한 데이터의 속성을 살펴보자.



나경제의 프로젝트

1 Kaggle에서 데이터 수집하여 속성 알아보기

① 데이터 가져오기



② 데이터 속성 확인하기

Kaggle에서 Credit Card Fraud Detection Data를 다운로드해서 압축을 푼다.

압축을 해제하면 'Credit Card' 데이터에서 31개의 속성에 대하여 284,807개의 레코드를 확인할 수 있다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF
1	Time	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	Amount	Class	
543	406	-2.31	1.952	-1.61	3.998	-0.52	-1.43	-2.54	1.392	-2.77	-2.77	3.202	-2.9	-0.6	-4.29	0.39	-1.14	-2.83	-0.02	0.417	0.127	0.517	-0.04	-0.47	0.32	0.045	0.178	0.261	-0.14	0	1	
625	472	-3.04	-3.16	1.088	2.289	1.36	-1.06	0.326	-0.07	-0.27	-0.84	-0.41	-0.5	0.677	-1.69	2.001	0.667	0.6	1.725	0.283	2.102	0.662	0.435	1.376	-0.29	0.28	-0.15	-0.25	0.036	529	1	
4922	4462	-2.3	1.759	-0.36	2.33	-0.82	-0.08	0.562	-0.4	-0.24	-1.53	2.033	-6.56	0.023	-1.47	-0.7	-2.28	-4.78	-2.62	-1.33	-0.43	-0.29	-0.93	0.173	-0.09	-0.16	-0.54	0.04	-0.15	239.93	1	
6110	6986	-4.4	1.358	-2.59	2.68	-1.13	-1.71	-3.5	-0.25	-0.25	-4.8	4.896	-10.9	0.184	-6.77	-0.01	-7.36	-12.6	-5.13	0.308	-0.17	0.574	0.177	-0.44	-0.05	0.252	-0.66	-0.83	0.85	59	1	
6331	7519	1.234	3.02	-4.3	4.733	3.624	-1.36	1.713	-0.5	-1.28	-2.45	2.101	-4.61	1.464	-6.08	-0.34	2.582	6.739	3.042	-2.72	0.009	-0.38	-0.7	-0.66	-1.63	1.489	0.567	-0.01	0.147	1	1	
6333	7526	0.008	4.138	-6.24	6.676	0.768	-3.35	-1.63	0.155	-2.8	-6.19	5.664	-9.85	-0.31	-10.7	-0.64	-2.04	-1.13	0.116	-1.93	0.488	0.365	-0.61	-0.54	0.129	1.488	0.508	0.736	0.514	1	1	
6336	7535	0.027	4.132	-6.56	6.349	1.33	-2.51	-1.69	0.303	-3.14	-6.05	6.755	-8.95	0.703	-10.7	-1.38	-1.64	-1.75	0.777	-1.33	0.588	0.371	-0.58	-0.67	-0.76	1.605	0.541	0.737	0.497	1	1	
6338	7543	0.33	3.713	-5.78	6.078	1.667	-2.42	-0.81	0.133	-2.21	-5.13	4.561	-8.87	-0.8	-9.18	-0.26	-0.87	1.313	0.774	-2.37	0.27	0.157	-0.65	-0.55	-0.72	1.416	0.555	0.531	0.404	1	1	
6340	7551	0.316	3.809	-5.62	6.047	1.554	-2.65	-0.75	0.056	-2.68	-4.96	6.439	-7.52	0.386	-9.25	-1.37	-0.5	0.784	1.494	-1.81	0.388	0.209	-0.51	-0.58	-0.22	1.475	0.491	0.519	0.403	1	1	
6429	7610	0.726	2.301	-5.33	4.008	-1.73	-1.73	-3.97	1.064	-0.49	-4.62	5.589	-7.15	1.68	-6.21	0.495	-3.6	-4.83	-0.65	2.25	0.505	0.59	0.11	0.601	-0.36	-1.84	0.352	0.595	0.099	1	1	
6448	7672	0.703	2.426	-5.23	4.417	-2.17	-2.67	-3.88	0.911	-0.17	-5.01	4.676	-8.17	0.639	-6.76	1.297	-3.81	-3.75	-1.05	1.554	0.423	0.551	-0.01	0.722	0.473	-1.96	0.319	0.6	0.129	1	1	
6474	7740	1.024	2.001	-4.77	3.819	-1.27	-1.73	-3.06	0.89	0.415	-3.96	3.572	-7.19	0.147	-5.25	1.678	-2.64	-1.31	-0.39	1.118	0.204	0.343	-0.05	0.71	-0.37	-2.03	0.367	0.395	0.02	1	1	
6531	7891	-1.59	3.262	-4.14	2.357	-1.41	-1.88	-3.51	1.516	-1.21	-6.23	5.451	-7.33	1.361	-6.61	-0.48	-2.6	-4.84	-0.55	0.352	0.316	0.502	-0.55	-0.08	-0.43	0.124	0.322	0.264	0.133	1	1	
6611	8090	-1.78	3.403	-3.82	2.625	-1.98	-2.73	-3.43	1.413	-0.78	-6.2	4.367	-8.24	0.346	-6.59	0.266	-3.03	-4.21	-1.21	-0.27	0.364	0.454	-0.58	0.046	0.462	0.044	0.306	0.531	0.244	1	1	

위 데이터는 실제 카드 거래 정보를 연구에서 사용하기 위하여 사용 내역을 연구자가 식별할 수 없도록 수치로 바꿔서 표현한 것이다.

카드 데이터의 속성 중에 V1~V28 속성에 해당되는 데이터는 실제 내용을 알아볼 수 없도록 수치로 바꾸어 표현되었으나, 동일한 데이터에 대해서는 동일한 수치로 표현되므로 기계학습을 위한 패턴을 찾아낼 수 있다.

이번 프로젝트에서는 Class 속성을 Label로 사용하게 된다.

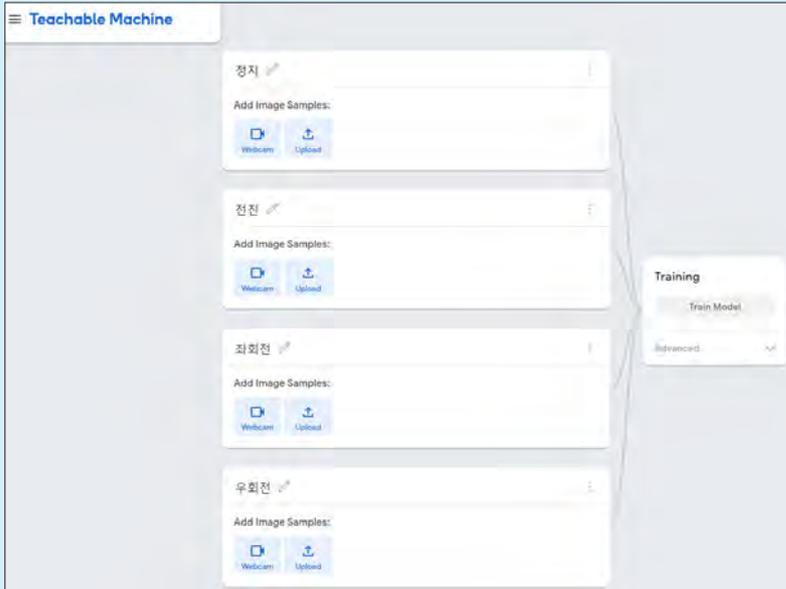
속성	의미
Time	카드 거래가 일어난 시점. 시간적 순서성만 가짐
V1~V28	통계적 방법을 이용하여 추출한 카드 거래 정보
Amount	거래 금액
Class	정상 사용 0, 부정 사용 1



2 Teachable Machine에서 데이터 분류하여 수집하기

1 데이터 분류하기

Teachable Machine에서 데이터를 직접 수집하기 위해서는 먼저 수집할 데이터를 분류해야 한다. 나우리의 프로젝트는 휠체어를 조종하기 위한 얼굴 방향 사진이 필요하므로 그림과 같이 정지, 전진, 좌회전, 우회전으로 클래스를 설정한다.

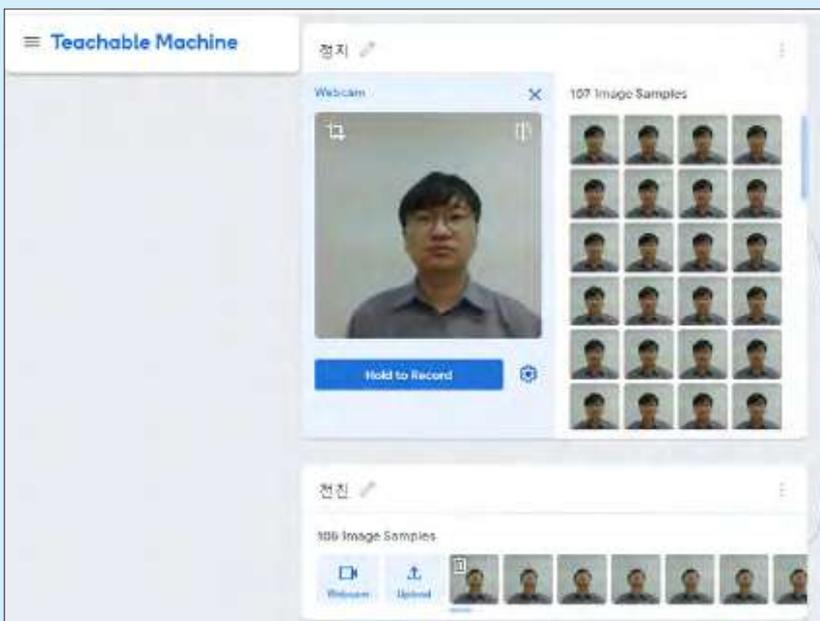


클래스는 데이터의 묶음으로 데이터를 분류하는 기준이에요.



2 데이터 수집하기

컴퓨터에 카메라가 장착되어 있어야 한다. 각 클래스의 이미지를 촬영할 때는 사진의 수는 많을수록 좋으며 최소 100장, 약 300장 정도의 사진을 수집한다. 사진은 **Hold to Record** 버튼을 누를 때마다 촬영되고, 버튼을 계속 누르고 있으면 연속 촬영된다.



클래스에서 사진 촬영이 끝난 뒤 잘못 촬영된 사진이 있으면 삭제해 주어야 합니다.



도움 자료

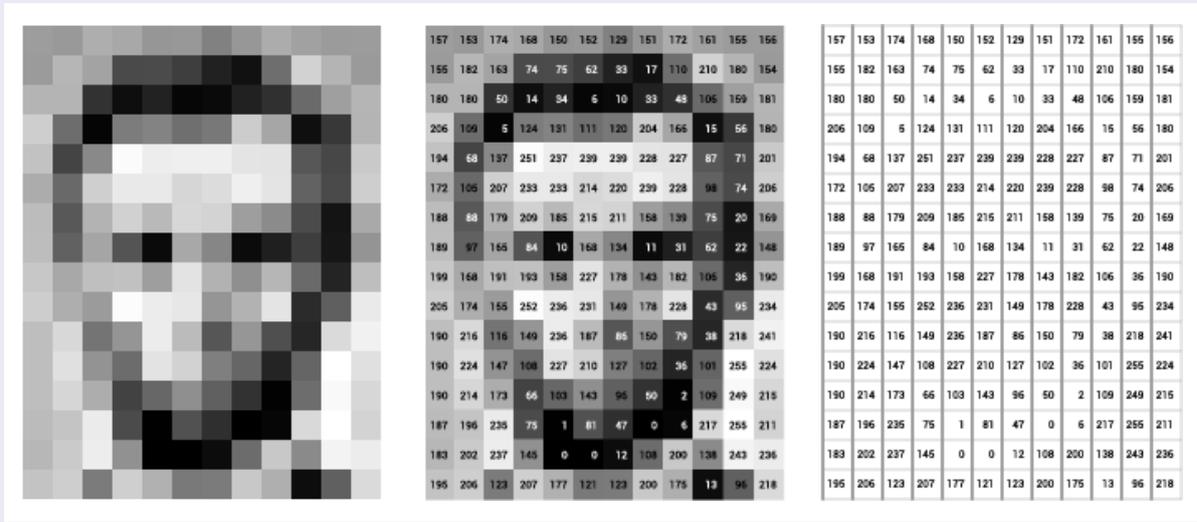
기계학습에서 사용하는 데이터

나경제의 프로젝트에서 사용한 데이터는 행과 열이 있는 정형 데이터이다. 암호화된 부분의 데이터는 비록 그 의미를 식별할 수 없더라도 인공지능 프로젝트에는 충분히 활용이 가능하다. 또 이런 정형 데이터는 내용을 확인할 때 열의 정보가 있기 때문에 어떤 내용인지 파악할 수도 있다.

그렇다면 나우리의 프로젝트에서 사용한 이미지는 어떤 형태일까? 프로젝트를 위해 데이터를 수집하였지만 어떻게 저장되어 있는지 쉽게 알 수 없다. 비정형 데이터인 이미지는 컴퓨터에서 아래 그림과 같이 픽셀이라는 작은 점으로 처리한다.

각각의 작은 점은 숫자를 이용하여 색상을 표시한다. 우리가 흔히 사용하는 흑백 사진의 경우에는 0~255의 값을 이용하여 색의 진하기를 표시하고, 컬러인 경우에는 빨강(R), 초록(G), 파랑(B) 각각의 값을 0~255의 값으로 표현하여 세 값의 농도로 색상을 표현한다.

인공지능 프로젝트에서 이미지를 사용하는 경우에는 이 픽셀 값을 사용하며 다른 정보는 사용하지 않는다.



4 정리하기

1 자신의 프로젝트를 진행하기 위해서 어떤 데이터가 필요하고 어디서 확보할 수 있을지 정리해 보자.

- 필요한 데이터의 종류

- 데이터를 확보할 수 있는 곳

2 기계학습에서 데이터의 역할이 무엇인지 설명해 보자.

- 데이터의 역할

3 나경제의 프로젝트에 필요한 데이터가 가지는 속성에는 어떤 것이 있는지 설명해 보자.

- 데이터의 속성의 종류

4 지금까지의 내용을 토대로 평가해 보자.

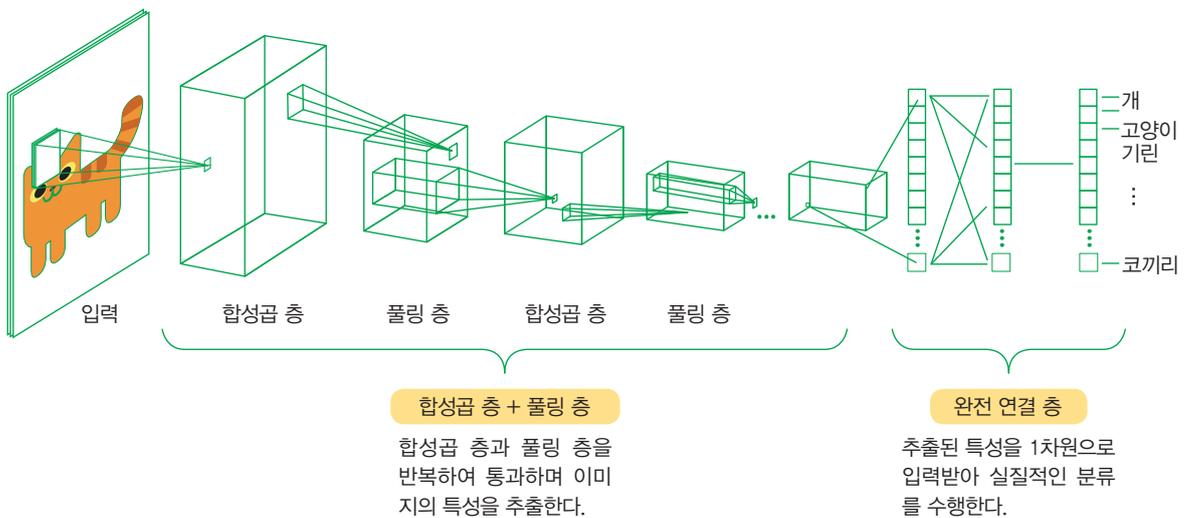
평가 요소	평가 항목	척도				
		1	2	3	4	5
데이터 수집과 속성	필요한 데이터를 수집하고 수집한 데이터의 속성을 파악할 수 있는가?					
데이터 종류	기계학습에 사용하는 정형 데이터와 비정형 데이터의 차이를 설명할 수 있는가?					
참여도	활동에 적극적으로 참여하였는가?					

합성곱 신경망 모델 (CNN: Convolutional Neural Network)

이미지 인식에 활용되는 Teachable Machine의 이미지 인식 모델은 기본적으로 합성곱 신경망 (CNN: Convolutional Neural Network)을 활용한다.

합성곱 신경망은 심층 신경망에 합성곱 층(Convolution layer)과 풀링 층(Pooling layer)이라는 고유의 구조를 더한 인공 신경망이다. 기존의 인공 신경망은 이미지 데이터를 단순히 숫자의 모음으로 인식하여 1차원의 숫자 배열 형태로 펼쳐 연산하였다. 반면, 합성곱 신경망에서는 이미지 데이터를 주입할 때, 이미지가 갖추고 있는 3차원 데이터 구조를 파괴하지 않고 마치 퍼즐들의 모음인 것처럼 전달하여 이미지의 특성을 잘 추출할 수 있도록 돕는다.

합성곱 신경망은 합성곱 연산과 풀링 연산을 반복하며 이미지의 특성을 추출해 나간다. 이는 크기가 큰 이미지 데이터를 신경망의 층을 통과하게 하여 데이터의 크기는 줄이고 이미지의 중요한 특성은 뽑아내는 것이 목적이다.



활동 4

나의 모델은 얼마나 똑똑해졌을까?

➔ 모델을 학습하고 성능을 평가할 수 있다.

1 문제 이해하기

다음 기사를 읽고, 성능 평가에 필요한 지표에 대해 알아보자.

2020년, 코로나19 진단 키트 시장에서 코로나19 검사 후 감염 여부 판정 속도가 가장 빠른 항원 진단 키트에 대한 수요가 폭증하고 있다. 기존에는 검사 후 감염 여부 판정까지 다소 시간이 소요되더라도 가장 정확도가 높은 유전자 증폭 방식의 분자진단(PCR) 검사가 코로나19 진단 키트 시장을 지배했다.



지난 11일 식약처에서 첫 국내 사용 코로나19 진단 키트 승인을 받은 에스디바이오센서 제품의 임상적 성능 시험을 진행한 결과, 정확도는 민감도 90%, 특이도 96%에 달했다. 민감도는 양성 검체를 양성으로 확진 판정하고, 특이도는 음성 샘플을 음성으로 판별할 수 있다는 의미로 둘 다 진단 키트 정확도를 평가하는 기준이다. 일각에서는 항원 진단 키트의 약점으로 바이러스 분비량이 적으면 항원 검사만으로 감염 여부를 완벽하게 판정하기 힘들다는 점을 지적한다. 하지만 코로나19 바이러스 농도는 감염 후 일주일까지 계속 높아졌다가 줄어들기 때문에 몸에 이상이 생겨 초기에 검사하면 정확도를 90% 이상으로 끌어올릴 수 있다는 게 업계 설명이다.

(후략)

[출처: 매일경제(<https://www.mk.co.kr/news/it/view/2020/11/1188015/>)]

위 기사에서 코로나 진단 키트의 성능을 평가한 기준은 무엇인지 생각해 보자.

🎯 무엇을 알아야 할까?

🔍 배경 지식

모델의 학습과 성능 평가

프로젝트를 위한 데이터를 확보했으면, 이 데이터를 이용하여 기계학습 모델을 학습 시킨다. 이때 제대로 된 모델이 만들어졌는가는 평가를 통해서 확인할 수 있다. 인공지능 모델을 평가할 수 있는 방법은 여러 가지가 있으며 그중 가장 많이 사용하는 방법은 정확도이다.

분류 모델에서 정확도는 정확히 분류된 샘플 수이다. 정확도를 구하는 식은 다음과 같다.

$$\text{정확도(accuracy)} = \frac{\text{정확하게 예측한 건수}}{\text{전체 예측 건수}}$$

정확도는 혼동 행렬(Confusion Matrix)을 활용하면 네 가지 형태로 성능을 평가할 수 있다.

		예측	
		Positive	Negative
실제	Positive	TP	FN
	Negative	FP	TN

- True Positive (TP) - 모델이 정답(Positive)을 맞힘.
- True Negative (TN) - 모델이 오답(Negative)을 맞힘.
- False Postive (FP) - 모델이 오답(Negative)을 정답(Positive)으로 잘못 예측함.
- False Negative (FN) - 모델이 정답(Positive)을 오답(Negative)으로 잘못 예측함.

예 나경제의 프로젝트에 위 혼동 행렬을 적용했을 때 TP와 TN을 합하여 정확도를 계산할 수 있다.

- TP (True Positive) - 올바른 카드 사용을 올바르다고 분류함.
- TN (True Negative) - 부정한 사용을 부정하다고 분류함.
- FP (False Postive) - 부정한 사용을 올바르다고 분류함.
- FN (False Negative) - 올바른 사용을 부정한 사용이라고 분류함.

2 모델 학습 플랫폼 이해하기

모델 학습을 위한 Orange3와 Teachable Machine의 사용 방법을 알아보자.

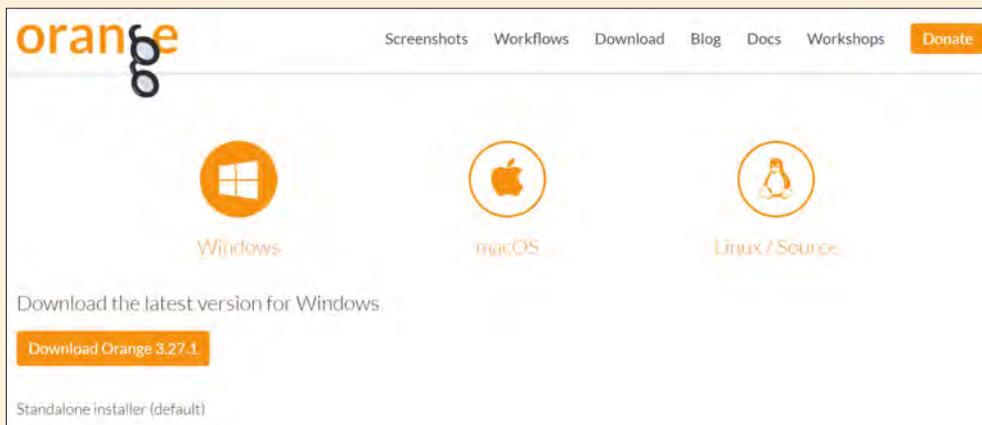


나경제의 프로젝트

1 인공지능 프로젝트를 위한 Orange3 설치하기

① 주소창에 URL을 입력하여 접속한다.

<https://orangedatamining.com/download/#windows>

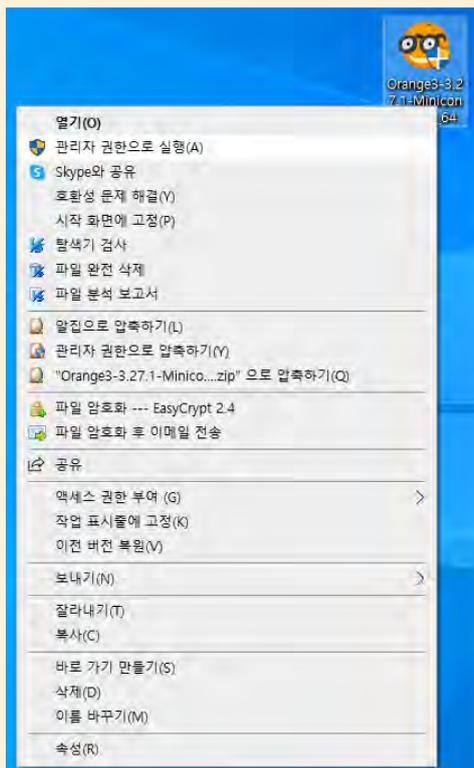


내 컴퓨터
운영체제에
해당하는 설치 파일을
다운로드하세요.



② 다운로드 버튼 **Download Orange 3.27.1** 을 눌러 다운로드를 실행한다.

③ 다운로드한 파일 아이콘을 관리자 권한으로 실행해 주면 설치가 진행된다.



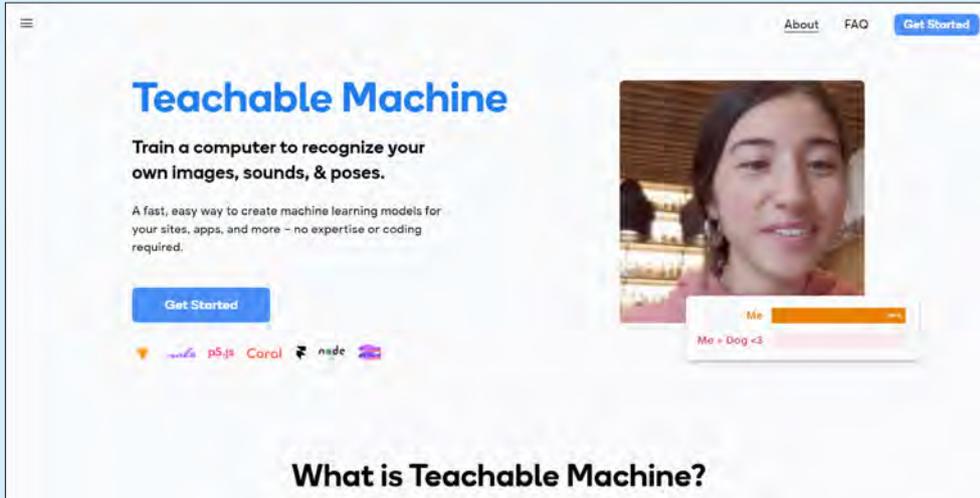


나우리의 프로젝트

2 인공지능 프로젝트를 위한 Teachable Machine 사용하기

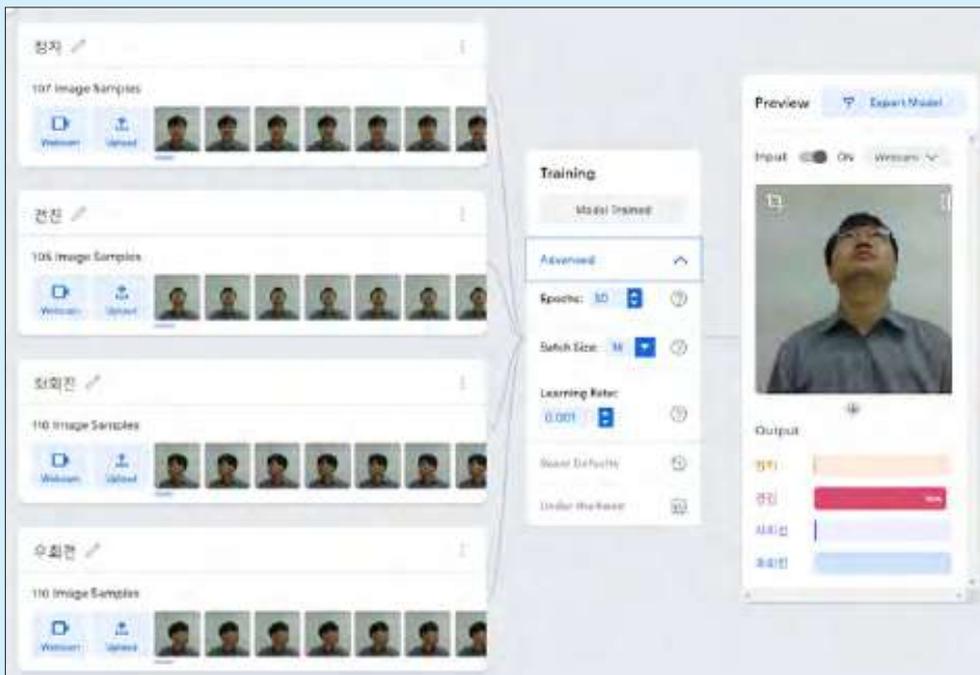
1 주소창에 URL을 입력하여 접속한다.

`https://teachablemachine.withgoogle.com/`



2 **Get Started** 버튼을 누른 뒤  (Image Project)를 선택한다.

3  (Webcam)을 이용하여 직접 사진을 찍어 정지, 전진, 좌회전, 우회전에 대한 학습 데이터를 수집한다.



4 **Train Model** 버튼을 이용하여 모델을 학습시킨다.

3 문제 해결하기

문제 해결에 필요한 학습 모델을 만들어 보자.



나경제의 프로젝트

1. orange3에서 학습 모델 만들기

1 훈련 데이터와 테스트 데이터 나누기

확보한 creditcard 데이터를 훈련 데이터와 테스트 데이터로 구분한다. 테스트 데이터는 훈련에 사용하지 않고 모델의 훈련이 끝난 뒤 성능의 평가를 위하여 사용한다. creditcard 데이터에는 정상 사용과 부정 사용을 나타내는 Class라는 속성이 있다. 정상 사용은 0, 부정 사용은 1의 값이 저장되어 있다.

creditcard 데이터에서 스프레드시트 프로그램을 이용하여 Class 속성이 0인 데이터 100개와, 1인 데이터 100개를 무작위로 추출하여 creditcard_test로 저장한다. 이때, creditcard_test의 파일은 쉼표(,)로 구분된 CSV 파일로 저장하도록 한다.

훈련 데이터는 모델의 훈련에 사용하고, 테스트 데이터는 학습한 모델의 실제 성능 측정에 사용한다.

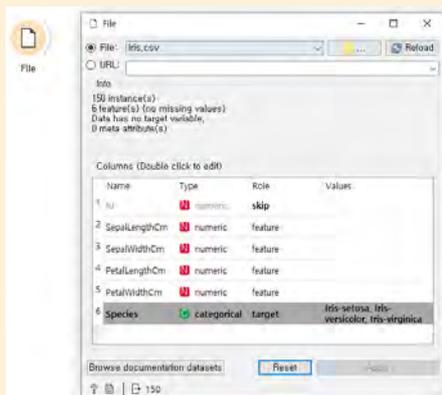
85	93742	-3.29	4.4	-8.39	4.45	-4.79	-4.24	-9.22	1.97	-2.91	-13.5	9.41	-18	-0.73	-15.4	0.35	-13.3	-21.3	-7.26	2.27	1.49	2.1	0.6	-0.33	0.45	0.59	-0.4	2.12	1.05	1	1
86	93823	-3.82	5.67	-9.24	8.25	-4.37	-3.45	-8.43	2.31	-5.34	-12	8.62	-18.7	-0.96	-15.3	-2.41	-12.1	-21.3	-8.05	0.16	1.12	1.99	0.08	-0.06	0.39	0.33	0.1	1.87	1	75.9	1
87	93824	-3.63	5.44	-9.14	10.3	-5.42	-2.86	-10.6	3.02	-4.89	-11.2	8.79	-18.6	-0.34	-15.6	-0.19	-12.4	-20.2	-6.89	2.59	1.35	2.31	0.98	-0.1	0.43	-0.44	0.65	1.69	0.86	8.54	1
88	93834	-3.77	5.89	-10.2	10.3	-5.61	-3.24	-10.6	3.27	-5.27	-11.2	8.88	-18.4	-0.23	-15	0.14	-12.2	-20.2	-7.05	2.5	1.19	2.25	0.55	0.38	0.38	-0.82	0.39	1.41	0.78	0.01	1
89	93853	-6.19	7.1	-13	8.01	-7.89	-3.97	-12.2	4.97	-4.25	-13	8.69	-17.2	0.07	-14.1	0.96	-12.4	-18.7	-6.52	3.52	0.48	2.5	0.48	0.48	0.36	-0.29	-0.2	-0.2	0.4	44.9	1
90	93853	-6.19	7.1	-13	8.01	-7.89	-3.97	-12.2	4.97	-4.25	-13	8.69	-17.2	0.07	-14.1	0.96	-12.4	-18.7	-6.52	3.52	0.48	2.5	0.48	0.48	0.36	-0.29	-0.2	-0.2	0.4	44.9	1
91	93853	-5.84	7.15	-12.8	7.03	-9.65	-2.94	-11.5	4.84	-3.49	-13.3	8.46	-17	0.1	-14.1	0.75	-12.7	-18.9	-6.63	4.01	0.06	2.46	1.05	0.53	0.47	-0.28	0.28	0.1	0.25	316	1
92	93853	-5.84	7.15	-12.8	7.03	-9.65	-2.94	-11.5	4.84	-3.49	-13.3	8.46	-17	0.1	-14.1	0.75	-12.7	-18.9	-6.63	4.01	0.06	2.46	1.05	0.53	0.47	-0.28	0.28	0.1	0.25	316	1
93	93856	-6.75	5.37	-10.1	9.06	-7.97	-2.26	-10.3	4.24	-5.32	-11.1	7.15	-17.2	-0.1	-11	-1.96	-12.4	-22.7	-9.26	1.98	-0.08	1.91	-0.35	0.43	0.67	-0.78	-0.25	1.16	0.2	210	1
94	93860	-10.9	6.73	-16.8	8.43	-10.3	-4.19	-14.1	7.17	-3.68	-15.2	8.03	-16.1	0.27	-15	-0.24	-11.9	-15.5	-5.75	4.13	-0.65	2.54	0.14	-1.02	0.41	0.11	-0.03	-1.46	-0.41	78	1
95	93860	-10.9	6.73	-16.8	8.43	-10.3	-4.19	-14.1	7.17	-3.68	-15.2	8.03	-16.1	0.27	-15	-0.24	-11.9	-15.5	-5.75	4.13	-0.65	2.54	0.14	-1.02	0.41	0.11	-0.03	-1.46	-0.41	78	1
96	93860	-10.6	7.25	-17.7	8.2	-10.2	-4.51	-13	6.78	-4.66	-14.9	8.39	-16.5	0.34	-14.2	0.56	-11.7	-15.8	-5.75	3.81	-0.81	2.72	0.7	-1.14	0.46	0.39	0.52	-1.42	-0.49	189	1
97	93860	-10.6	7.25	-17.7	8.2	-10.2	-4.51	-13	6.78	-4.66	-14.9	8.39	-16.5	0.34	-14.2	0.56	-11.7	-15.8	-5.75	3.81	-0.81	2.72	0.7	-1.14	0.46	0.39	0.52	-1.42	-0.49	189	1
98	93879	-13.1	7.35	-18.3	10.6	-11.7	-3.66	-14.9	8.81	-5.42	-13.2	6.36	-15.5	0.66	-11.4	-2.45	-9.83	-18.2	-7.27	0.62	-1.38	2.76	-0.27	-0.41	0.52	-0.74	-0.17	-2.5	-0.71	30.3	1
99	93879	-13.1	7.35	-18.3	10.6	-11.7	-3.66	-14.9	8.81	-5.42	-13.2	6.36	-15.5	0.66	-11.4	-2.45	-9.83	-18.2	-7.27	0.62	-1.38	2.76	-0.27	-0.41	0.52	-0.74	-0.17	-2.5	-0.71	30.3	1
00	93879	-12.8	7.51	-20.5	7.47	-11.6	-5.14	-14	8.33	-4.34	-15.6	7.61	-15.6	0.5	-13.2	-0.8	-10.3	-16	-5.88	1.57	-1.38	2.97	0.62	-0.77	0.43	-0.1	-0.01	-2.4	-0.72	104	1
01	40149	-1.05	-0.35	0.69	0.75	-1.02	0.48	2.21	-0.07	-0.9	-0.59	0.73	0.03	-0.53	0.55	0.14	0.48	-0.83	0.95	-0.03	0.92	0.44	0.39	0.79	-0.06	0.42	-0.33	-0.13	0.05	468	0
02	40149	-1.66	-0.97	1.73	-0.48	-2.29	0.98	-0.56	0.76	-0.37	0.12	0.02	-0.42	-0.8	-0.62	-1.14	0.98	0.77	-0.18	2.07	0.33	0.29	0.89	0.04	0	0.1	0.03	0.25	0.02	267	0
03	40149	1.17	-1.29	0.88	-0.59	-1.74	-0.02	-1.34	0.32	0.09	0.72	0.35	-1.4	-2.62	0.05	0.01	1.62	0.11	-0.05	0.87	0.09	0.36	0.74	-0.2	-0.01	0.34	-0.04	0.01	0.02	95	0
04	40149	-1.88	-5.98	0.01	0.39	-3.29	1.28	0.5	0.14	2.94	-2.27	-0.49	1.33	-0.33	-0.71	-0	-1.57	1.33	-0.62	-0.14	2.91	1.07	0.3	-1.42	0.29	-0.2	0.13	-0.16	0.29	1536	0
05	40150	-1.59	1.07	-0.12	0.21	-1.74	0.07	2.56	-0.33	-0.52	-0.55	-0.49	0.62	1.3	0.07	0.15	0.41	-0.77	-0.21	-0.97	-0.64	0.04	0.64	-0.25	0.47	-0.42	-0.69	-0.43	-0.57	436	0

주의할 점

creditcard 데이터에서 테스트를 위해 추출한 데이터는 삭제해 주어야 한다. 즉, 훈련 데이터에 테스트 데이터를 포함하지 않도록 해야 진정한 테스트가 된다.

2 훈련 데이터 불러오기

1 Orange3의 Data 위젯 그룹에서 File 위젯을 가져와서 더블클릭한다.



File 위젯에 마지막으로 사용한 파일의 데이터가 나타나요.



② 파일 열기 버튼을 이용하여 미리 준비해 둔 훈련용 데이터 파일을 가져온다.

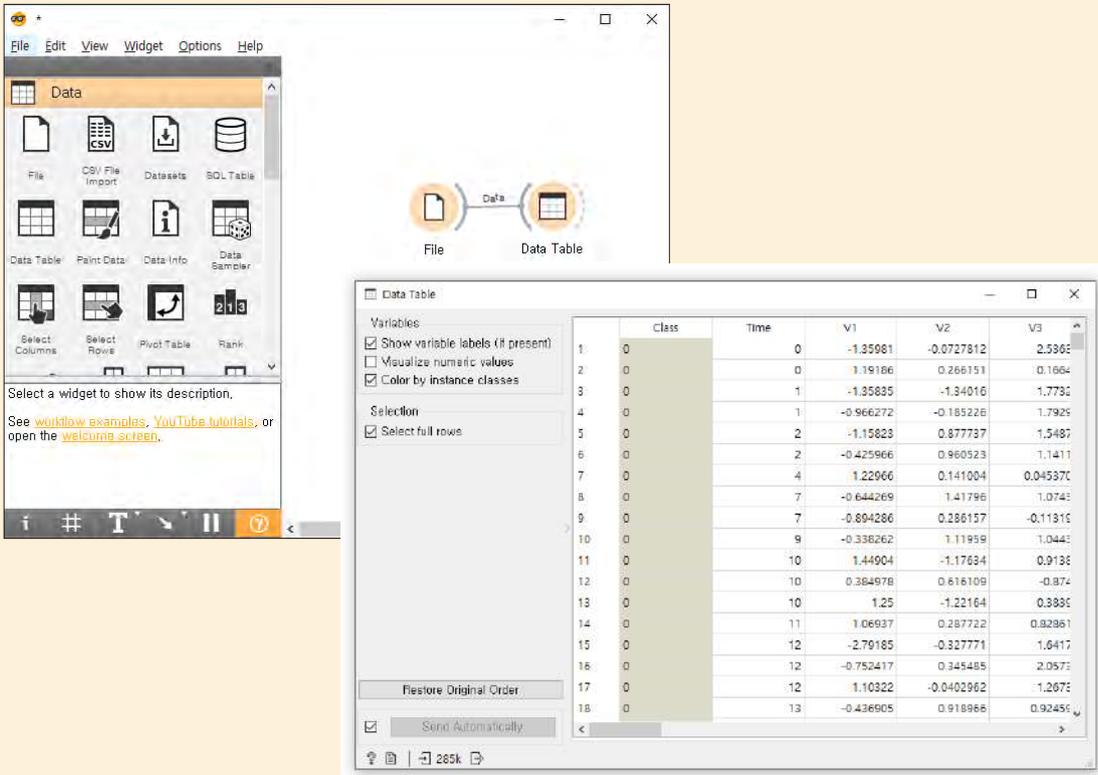
③ 훈련 데이터의 속성을 확인한다.

신용카드 부정 사용 예측 모델에서는 부정 사용인가, 정상 사용인가를 예측할 것이므로 예측할 값이 담겨있는 Class속성을 target으로 변경해 준다. Orange3에서는 예측을 위해 사용할 속성은 feature로, 예측할 값은 target으로 설정한다.



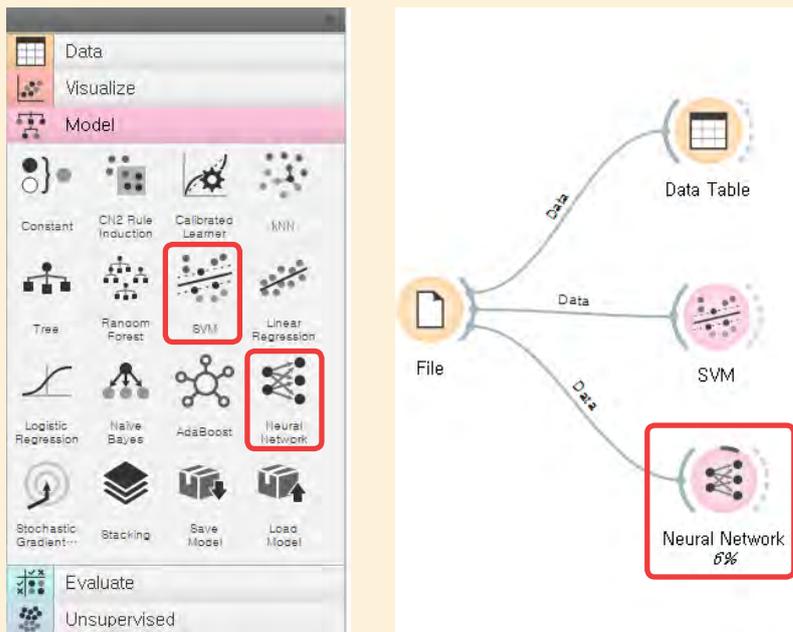
Name	Type	Role	Values
V22	numeric	feature	
V23	numeric	feature	
V24	numeric	feature	
V25	numeric	feature	
V26	numeric	feature	
V27	numeric	feature	
V28	numeric	feature	
Amount	numeric	feature	
Class	categorical	feature	0, 1

④ Data Table 위젯을 활용하여 훈련 데이터 내용을 확인한다. Orange3는 연결선을 연결하면 자동으로 위젯이 실행된다. 보기 편하게 실제 위치와는 달리 Target이 제일 처음에 표시된다.



3 학습 모델 선정하고 학습시키기

① 모델 그룹에 가면 다양한 기계학습 모델에 해당하는 위젯을 볼 수 있다. 여기서 SVM과, Neural Network를 선택하고 File 위젯을 그림과 같이 연결해 준다.



Orange3에서 사용할 다양한 분류 모델 중 여기서는 Support Vector Machine, Neural Network 2개의 모델을 사용해 보고 각각 비교해 봅시다.



② 위 그림에서와 같이 별도의 실행 명령을 주지 않아도 Neural Network가 동작하여 훈련 데이터를 이용하여 현재 6% 정도 훈련하고 있는 모습을 볼 수 있다.

도움 자료

SVM(Support Vector Machine)과 NN(Neural Network)

• SVM(Support Vector Machine)

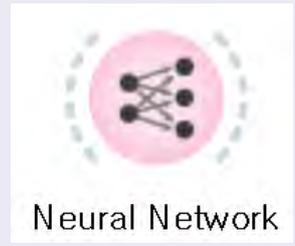
SVM은 분류에 많이 사용하는 기계학습 모델이다. SVM의 가장 큰 특징은 단순히 두 그룹을 나누는 구분선이 아니라 가장 좋은 구분선(hyperplain)을 찾는 것이다.

1996년 블라디미르 베프닉(Vladimir N. Vapnik)이 개발한 이후 글자 인식 분야에서 그 성능을 인정받아 기계학습 분야에서 분류를 위한 대표적인 알고리즘이 되었다.



• NN(Neural Network)

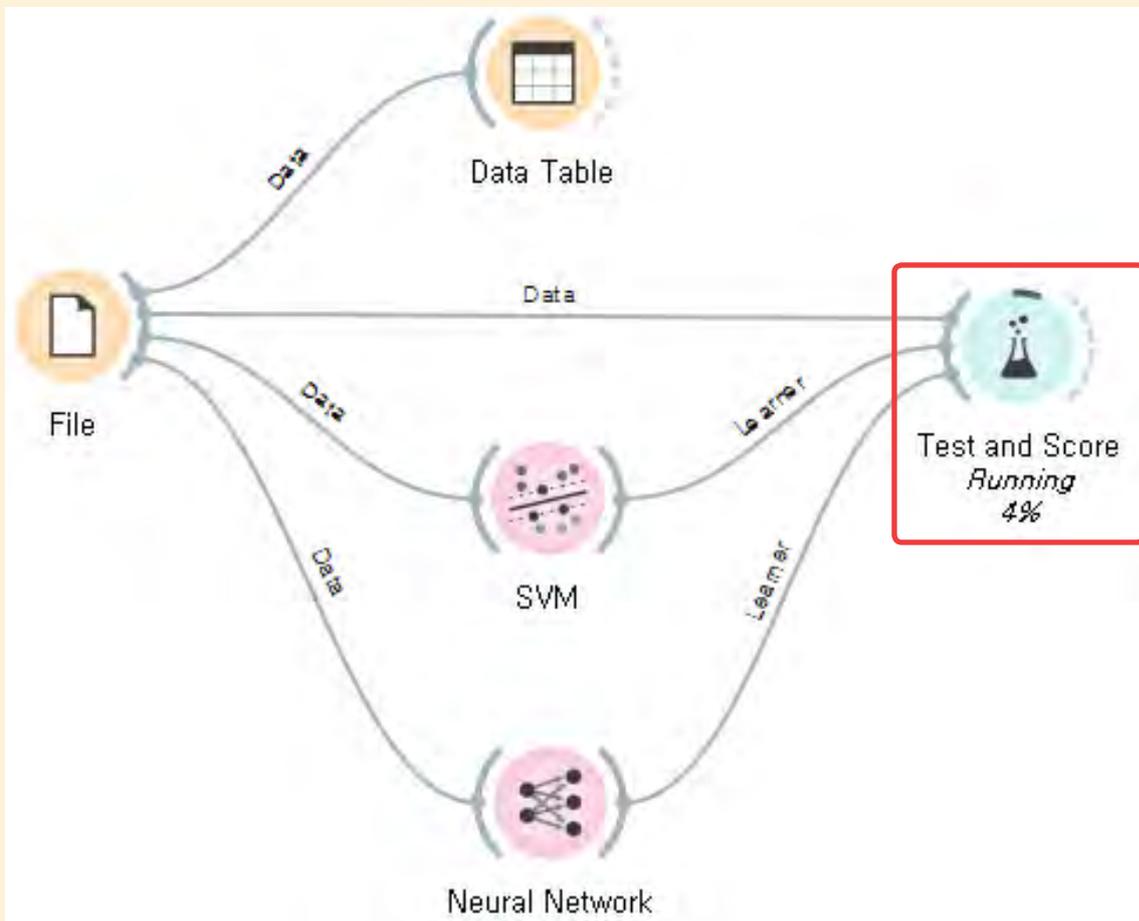
NN(Neural Network)은 일반적으로 DNN(Deep Neural Network)라고도 부른다. NN은 동물의 뉴런(신경세포)를 모방한 퍼셉트론(perceptron)을 기반으로 한다. 분류뿐 아니라 예측 등 다양한 기계학습 분야에 활용할 수 있다. 퍼셉트론 하나로는 해결할 수 있는 문제가 많지 않으나 퍼셉트론 여러 개를 순차적, 계층적으로 연결하여 NN을 만들고 데이터로 훈련하면 다양한 문제를 해결하는 데 활용할 수 있다.



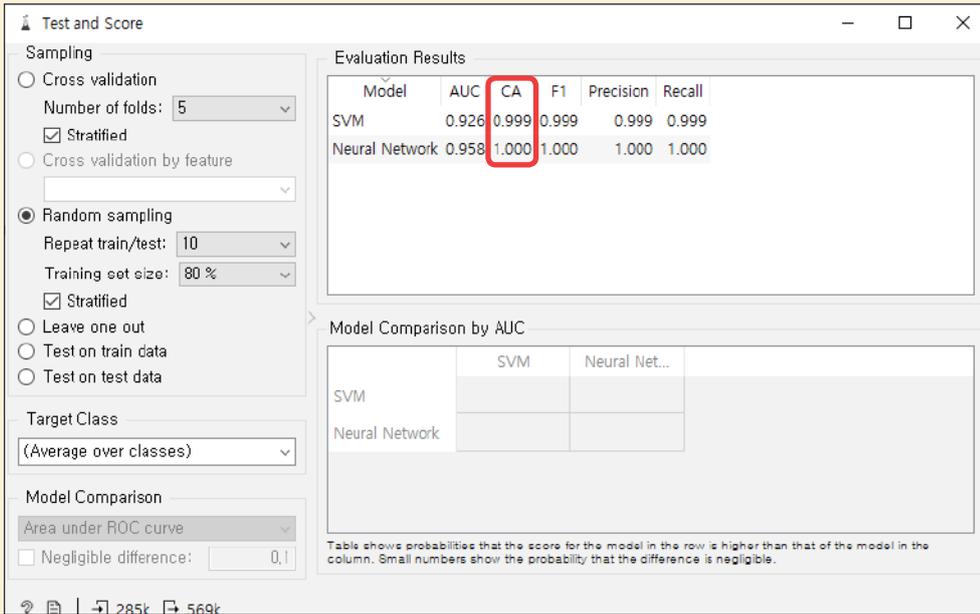
4 학습 결과 확인하기

① Evaluate 그룹의 Test and Score 위젯과 각 분류 모델, 훈련 데이터를 연결한다.

Test and Score 위젯을 추가한 다음, 각 분류 모델을 그림과 같이 연결하고 훈련 데이터로 사용한 데이터도 연결해 준다.



② Test and Score 위젯을 더블클릭하여 각 모델의 성능을 확인한다.



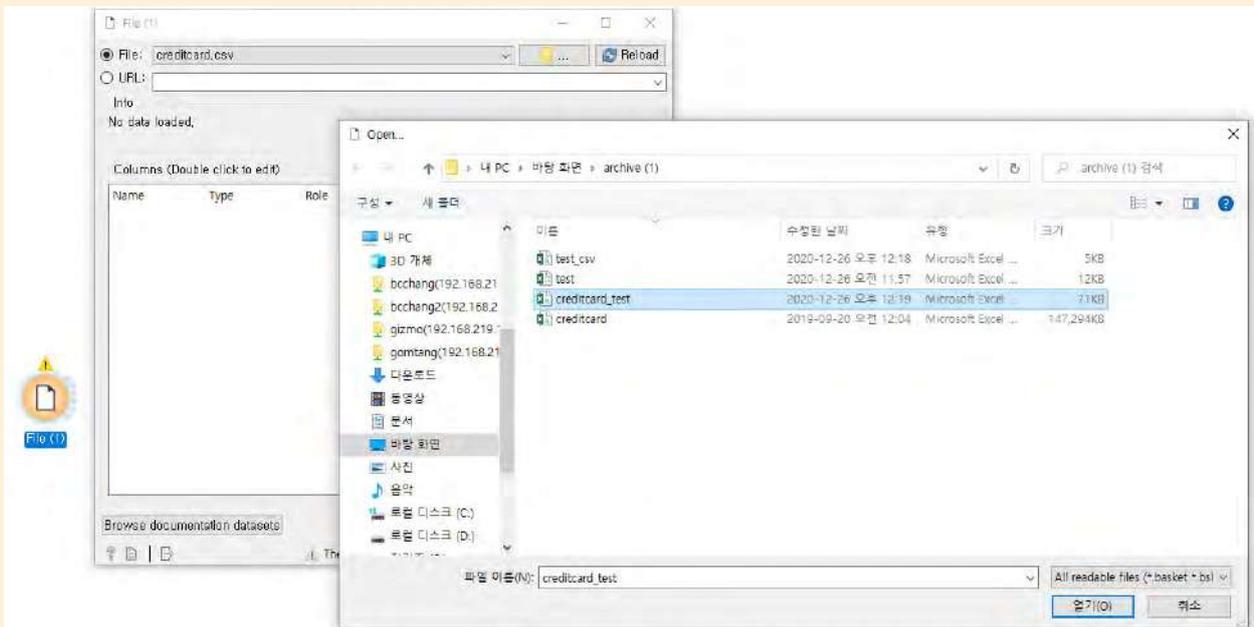
▲ Test and Score 결과 화면

CA가 우리가 사용할 Accuracy인데 살펴보면 여기서 사용한 기계학습 모델 대부분이 99%의 정확도를 보이고 있음을 알 수 있다. 즉, 모델이 신용카드 데이터에 대하여 정상적인 사용과 부정적 사용을 잘 찾고 있음을 알 수 있다.

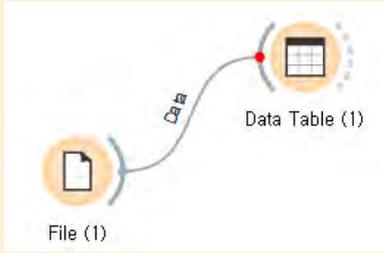
5 테스트 데이터를 이용한 성능 측정하기

앞서 실행한 Test and Score 위젯은 훈련 데이터를 이용하여 정확도를 측정한 것이다. 그러나 실제로 활용하려면 훈련에 사용하지 않은 데이터를 입력하였을 때, 얼마나 정확한지를 측정하는 것이 필요하다. 처음 시작할 때 훈련 데이터와 테스트 데이터를 구분하였는데 이때 떼어 놓은 테스트 데이터를 이용하여 성능을 측정해 보자.

① File 위젯을 이용하여 테스트 데이터를 불러온다.



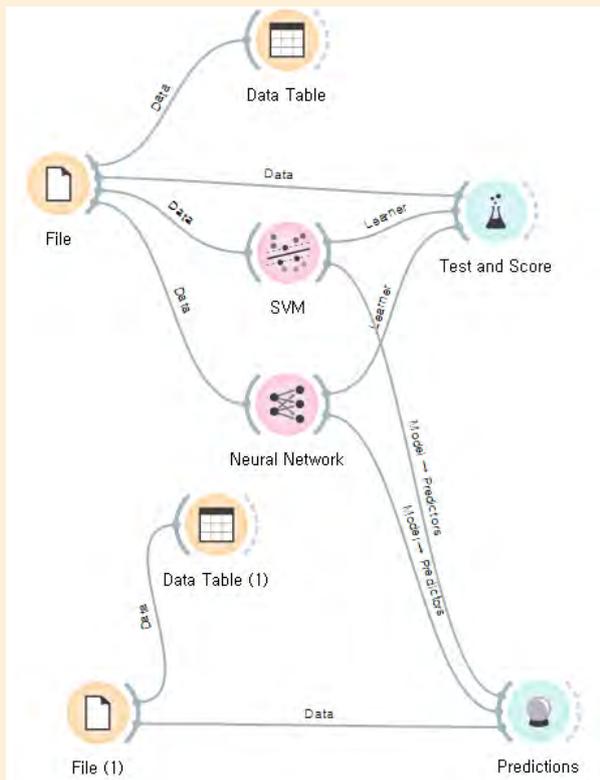
② Data Table 위젯을 이용하여 데이터 내용을 확인한다.



	Time	V1	V2	V3	V4	
1	4.0	1.22965763450...	0.14100350704...	0.04537077358...	1.20261273673...	0.1
2	7.0	-0.6442694423...	1.41796354547...	1.07438037635...	-0.4921990184...	0.9
3	7.0	-0.8942860822...	0.28615719627...	-0.1131922127...	-0.2715261300...	2.6
4	39.0	-1.3308815753...	0.26754022707...	-0.1684695735...	-0.7012321479...	3.2
5	40.0	1.11069200372...	0.08194190138...	0.40606303731...	1.29074038305...	-0.1
6	41.0	1.15431211678...	0.26546231732...	0.38487052407...	0.57500704018...	-0.2
7	41.0	0.98606332555...	-0.2029645867...	-0.4927681473...	0.40769127471...	0.3
8	41.0	1.13875933590...	-1.1929528274...	1.40713083554...	-0.3300701033...	-2.0
9	41.0	1.14552438734...	0.57506796812...	0.19400845998...	2.59819174498...	-0.0
10	42.0	-0.2493613279...	0.39922737286...	0.06800933370...	-1.0606188263...	2.4
11	7610.0	0.72564573981...	2.30089443776...	-5.3299761830...	4.00768280468...	-1.7
12	7672.0	0.70270990009...	2.42643280600...	-5.2345132958...	4.41666124290...	-2.1
13	34256.0	0.53927591389...	1.55488976307...	-2.0661803332...	3.24161746276...	0.1
14	34521.0	1.08123377540...	0.41641393604...	0.86291874065...	2.52086280678...	-0.0
15	34634.0	0.33349949068...	1.69987344643...	-2.5965611669...	3.64394526899...	-0.1
16	17230.0	-0.4693267051...	1.11145278248...	2.04100272504...	1.73159516277...	0.1
17	17520.0	-5.2680532208...	9.06761342731...	-15.960728134...	10.2966027898...	-4.7
18	17838.0	-5.1878781083...	6.96770866359...	-13.510931139...	8.61789514115...	-1.1

③ Prediction 위젯을 이용하여 모델 성능을 평가한다.

- 앞서 훈련한 SVM, Neural Network 모델을 Prediction 위젯의 입력으로 사용한다.
- 이번에 사용한 데이터는 새로운 테스트 데이터이므로 그림과 같이 새 데이터를 연결해 준다.



④ 테스트 데이터에 대한 모델의 성능을 측정한다.

각 모델의 결과를 살펴볼 수 있는데, 파란색과 빨간색으로 표시된 것은 예측값이 각각 0과 1일 확률을 의미한다.

Model	AUC	CA	F1	Precision	Recall
SVM	0.97	0.695	0.625	0.796	0.695
Neural Network	0.84	0.802	0.781	0.850	0.802

⑤ 각 모델의 테스트 데이터에 대한 정확도 측정하기

결과표를 이용하여 각 모델의 정확도를 찾아 표에 써 보자.

모델	정확도
Support Vector Machine	
Neural Network	

⑥ 각 모델의 성능을 보고 모델을 선택해 보자.

나의 프로젝트에 적합한 모델을 선택해 보고 그 이유를 적어보자.

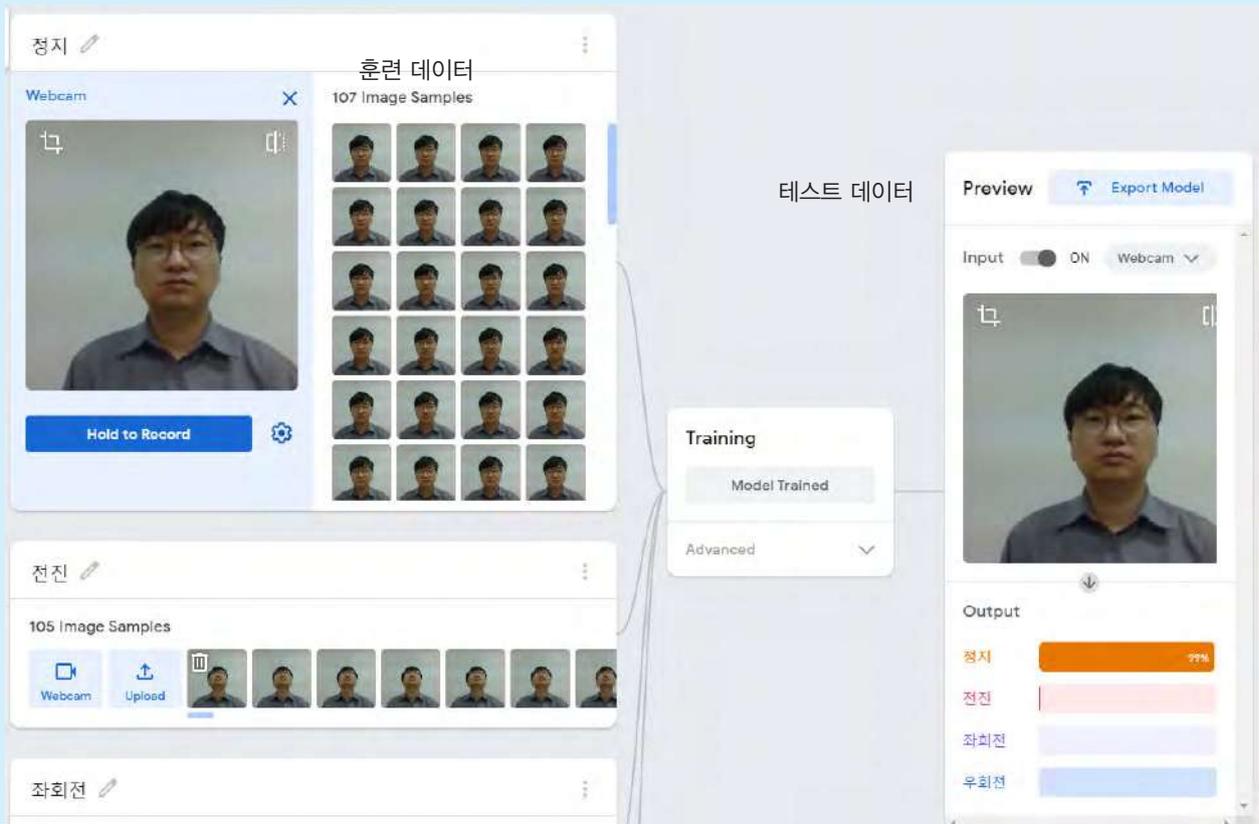


나우리의 프로젝트

2. Teachable Machine에서 학습 모델 만들기

1 훈련 데이터와 테스트 데이터 나누기

나우리의 프로젝트는 자신의 얼굴로 학습시키고 제대로 학습했는지를 확인하는 과정이 모두 Teachable Machine에서 실시간 웹캠으로 이루어진다. 이 경우는 별도의 테스트 데이터의 구분 없이 모델의 훈련 후, 실시간 창에서 모델의 정확도를 측정한다. 다만 모델의 훈련과 평가 시에 모두 저장된 데이터를 사용할 수 있다.



도움 자료

정확도의 함정

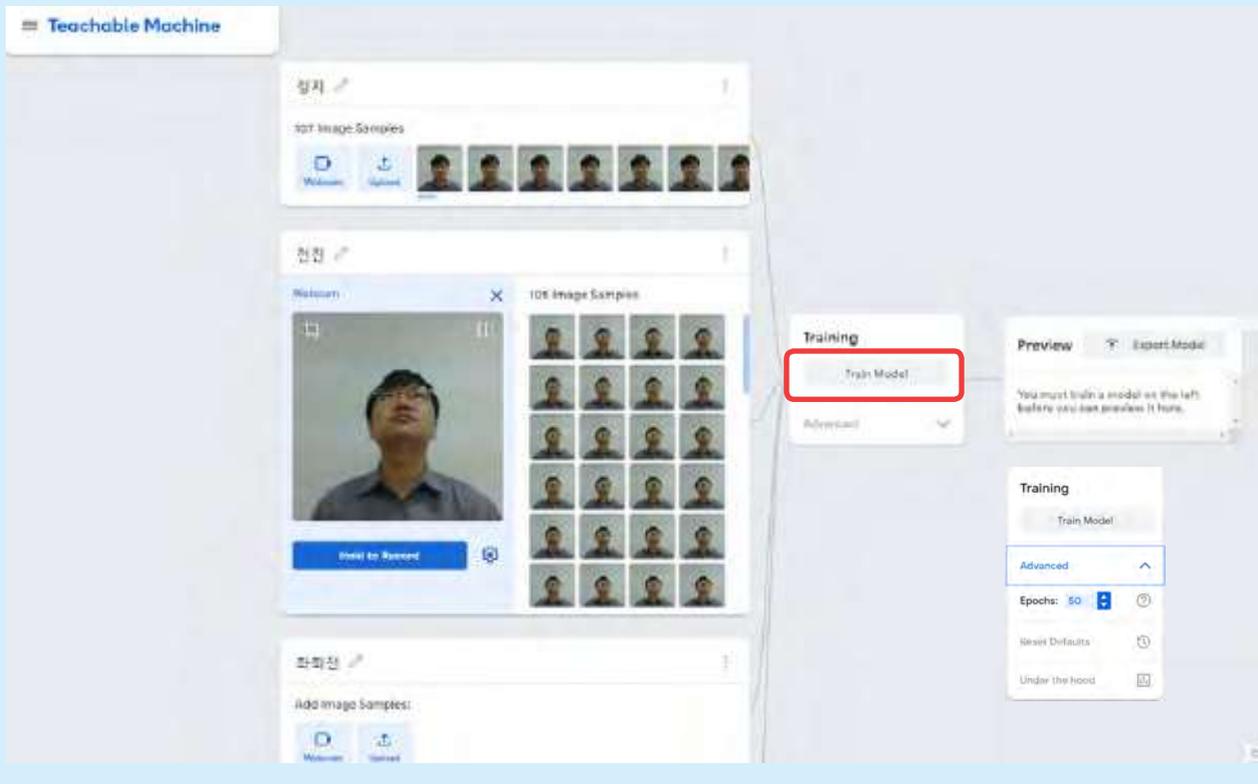
정확도가 높다고 모두 좋은 모델은 아니다. 예를 들어, 산불을 자동으로 감지하는 인공지능 시스템이 A와 B가 있다고 가정하자. 이 두 시스템은 총 200건의 임무에 투입되었는데 그중에 실제 산불인 경우가 100건, 산불이 아닌 경우가 100건 있었다.

- A 시스템: 산불이 아닌 경우 100건은 모두 산불이 아니라고 했고, 산불인 경우 100건 중 2건은 산불이 아니라고 했다.
- B 시스템: 산불이 아닌 경우 100건 중 2건은 산불이라고 했고, 산불인 경우 100건 모두를 산불로 감지했다.

두 시스템 모두 정확도는 98%가 된다. 하지만 산불 감시라는 임무를 감안하면 B 시스템이 더 적합함을 알 수 있다. 정확도 외에 모델의 성능을 평가를 위한 다양한 지표(정밀도, 재현율, F1-score, 민감도, 특이도)가 존재한다.

2 데이터 수집 후 Training 버튼으로 학습시키기

Training은 수집한 사진 자료로 이루어지며 Teachable Machine의 이미지 인식 모델은 기본적으로 CNN 모델을 활용한다.



도움 자료

알아 두면 도움되는 인공지능 학습 용어(Epoch, Batch size, Learning Rate)

• Epoch(세대)

전체 데이터 셋에 대한 반복 학습 횟수. 즉 50번의 Epoch는 전체 데이터 셋에 대해 50번의 학습을 완료한 것이다. Epoch 값을 너무 작게 설정하면 학습을 온전하게 하지 않아 모델의 성능이 좋지 않고 값을 너무 크게 잡으면 훈련 데이터에 대해서만 특화되어 다른 데이터에 대해서 성능이 좋지 못하다. Epoch가 너무 커도 테스트 성능이 좋지 않다.

• Batch size(배치 사이즈)

전체 데이터 셋을 batch size 크기의 작은 크기로 나누어 학습한다. batch size가 5라면 5개의 샘플 데이터를 학습하고 학습 모델이 갱신된다. 너무 작게 설정하면 학습 모델이 자주 갱신되어 학습 시간이 길게 걸리고 너무 크게 설정하면 자료에 따라, 특히 자료가 편향되면 학습에 반영되는 것이 둔하여 학습 결과 성능이 좋지 않다.

• Learning Rate(학습률)

반복 학습 시 경사 하강법(Gradient Descent)라는 학습 방법을 사용하는데, 이 알고리즘의 매개 변수가 Learning Rate이다. 너무 작으면 여러 번 반복에도 최적의 성능으로 모델이 학습되지 못하고 너무 크면 최적의 상태를 지나쳐 버리는 문제가 발생한다.

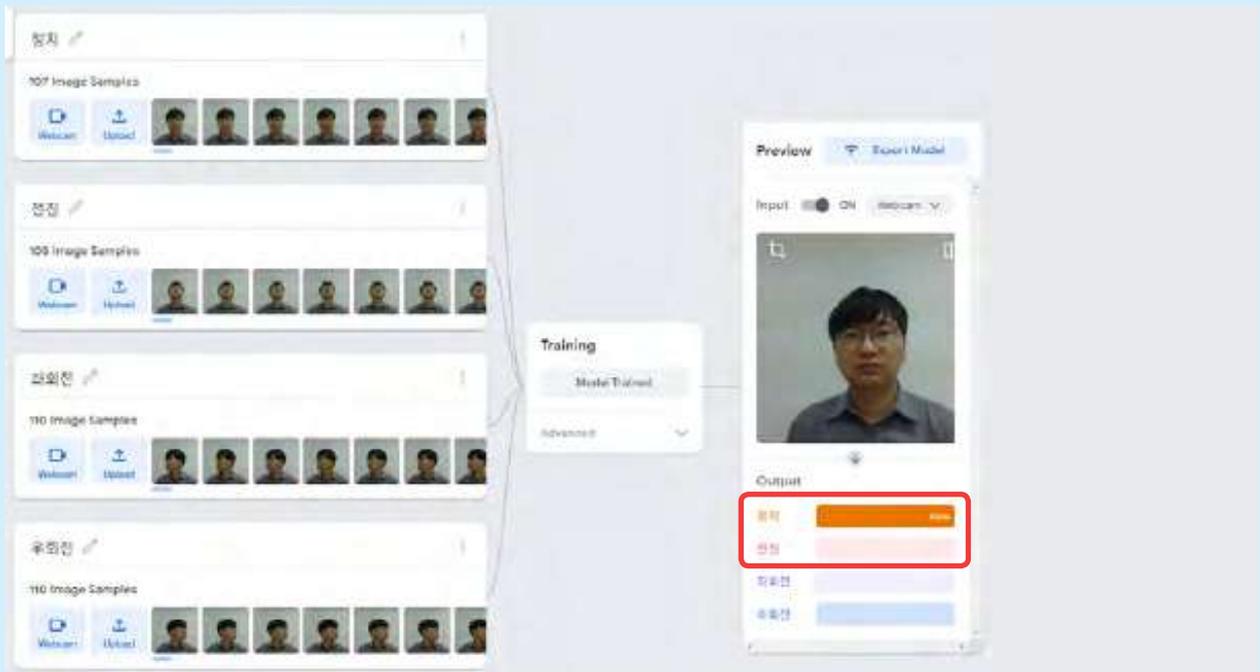
3 학습 결과 확인하기

이미지 분류 결과를 확인한다. Teachable Machine은 결과를 신뢰도로 나타낸다.

아래 결과처럼 teachable machine은 결과를 각 분류 항목에 대항하는 확률, 즉 신뢰도로 나타낸다. 그림을 보면 정지가 97%, 전진이 3%의 확률을 보이므로 이는 정지로 판단할 수 있다.

기계학습의 분류 모델은 일반적으로 2개 이상의 항목에 대하여 판단할 때는 각 항목의 확률 즉, 신뢰도를 결과로 제공한다. 신뢰도의 합은 1로 즉, 100%이다. 이것을 정확도의 형태로 바꾸려면 일정값 이상이면 해당 항목으로 분류한다고 정하고 정확도를 계산한다.

예를 들어 90% 이상이면 해당 값으로 분류한다고 정하는 경우에는 위 그림에서 정지가 97%이므로 이 사진을 정지로 분류한 것으로 볼 수 있고, 만약, 사진이 정지에 해당하면 정답, 그렇지 않으면 오답이 된다.



4 Teachable Machine에서 이미지 데이터 사용 계획 세우기

나의 프로젝트가 이미지 데이터를 사용한다면, Teachable Machine에서 분류 항목(클래스)을 어떻게 만들지, 자료는 어떻게 수집할지 등에 대해 생각해 보자.

4 데이터 양에 따른 성능 실험하기



나경제의 프로젝트

- 1 creditcard 데이터는 284,807개의 레코드를 가지고 있다. 해당 데이터에서 부정 거래를 한 데이터는 492개이다. 정상 데이터 284,315중 약 500개만 무작위 표집하여 학습에 사용하고 기존 모델과 성능을 비교해 보자. 성능이 같을까?



나우리의 프로젝트

- 2 정지, 전진, 좌회전, 우회전 데이터를 각각 100개 정도만 활용하였다. 이번 실험에서는 각 항목에 대하여 300개씩 촬영하여 실험해 보자. 성능이 같을까?

5 정리하기

- 1 데이터가 많으면 항상 모델의 성능이 좋아질까? 내 생각을 써 보자.

•

•

•

•

•

- 2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
테스트 데이터	테스트 데이터의 역할을 설명할 수 있는가?					
성능 평가	기계학습 모델의 성능을 평가 지표를 설명할 수 있는가?					
참여도	활동에 적극적으로 참여하였는가?					



기계학습 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

- 1 내가 다니는 학교, 내가 사는 마을, 우리나라, 더 나아가 세상이 가지고 있는 인공지능으로 해결할 수 있는 문제를 찾아보고 그 문제를 4Ws를 이용하여 구체화해 보자.

who	누구	
what	무슨 문제	
where	어떤 맥락/상황에서	
why	해결 이유/목적	

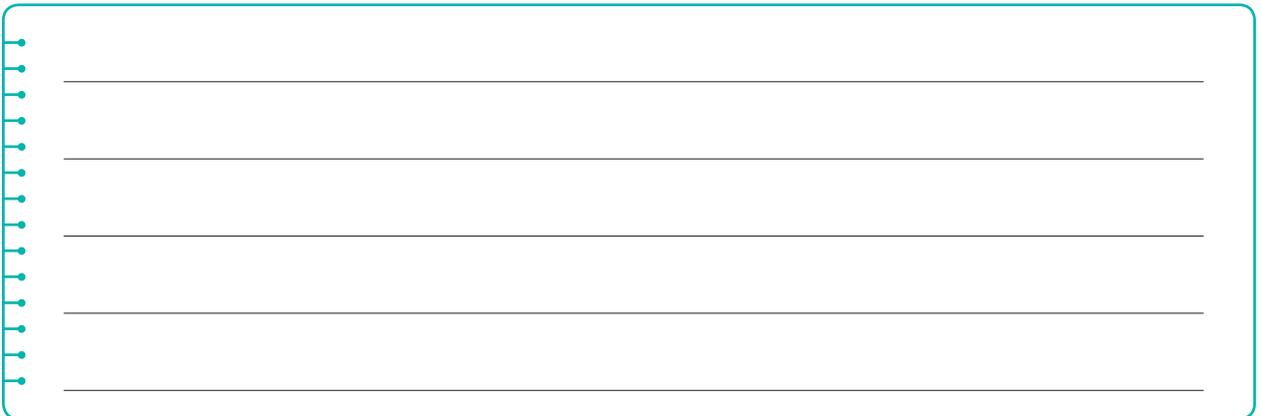
- 2 문제를 찾았다면 관련 있는 데이터를 어디서 획득할 것인지 제안해 보자. 그리고 획득한 데이터가 어떤 특징을 가지게 될지 예측해 보자.

- 3 문제 해결을 위해 적합한 모델을 찾아보자.

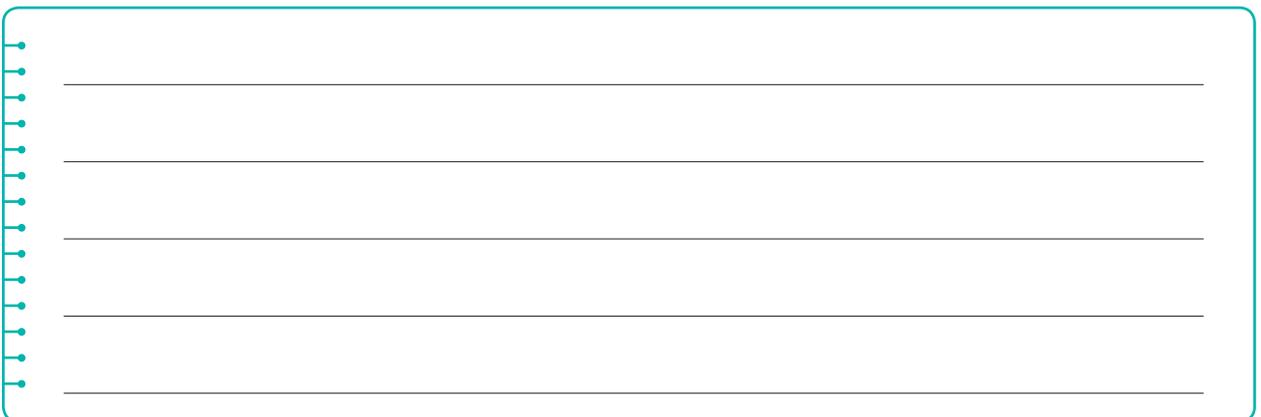
4 3 에서 만든 모델을 학습시킨 뒤 평가해 보고, 원하는 성능이 나오지 않는다면 그 이유를 파악해 보자.



5 모델을 공유하고, 세상에 이로움을 주기 위한 방법을 구체적으로 설계해 보자.



6 자신의 프로젝트를 발표하고, 서로 공유해 보자.



V 인공지능 윤리

1 단계 핵심 개념 알기

- 데이터 편향성
- 윤리적 딜레마



2 단계 창의융합적 사고로 해결하기

- ▶ 1 편향성은 어떻게 생기는 걸까?
- ▶ 2 인공지능의 선택은? 당신의 선택은?



3 단계 공감 나누기

- '인간적 기술 센터' 체험하기

③ 인공지능 윤리 단원에서는 무엇을 배울까요?

데이터 편향성과 인공지능 딜레마를 체험해 보고, 생활 속 문제 해결을 토대로 인공지능 윤리의 필요성을 이해합니다.





1단계 핵심 개념 알기

인공지능의 윤리적 이슈

✔ 인공지능은 가치중립적일까? 데이터 편향성과 딜레마를 알아보자.

인공지능과 같이 새롭게 등장하는 기술은 기존의 규범이나 윤리적 가치와 갈등을 일으키며 혼란스러운 상황을 만들 수 있다. 그렇기 때문에 새로운 기술의 영향력과 올바른 설계 방법에 대해 미리 고민하는 과정이 필요하다. 본 단원에서는 인공지능의 데이터 편향성과 윤리적 딜레마를 살펴보고, 윤리적 인공지능 설계의 필요성과 방법에 대해 알아본다.

1 데이터 편향성

Q 개념 탐색 인공지능은 모두에게 공정할까?

인공지능의 성능은 기계학습 모델을 학습시키는 데 사용하는 데이터에 의해 결정되는 경우가 많다. 하지만, 데이터를 수집하고 학습시키는 과정에서 인종, 성별, 장애, 종교 등과 관련해 인간이 의도하지 않았던 편견이 반영될 가능성이 있다.

데이터 편향성에 대한 개념과 발생 원리를 알아보고, 이를 해소하기 위한 방안을 탐색한다.

1 데이터 편향성

데이터 편향성이란, 기계학습 모델을 학습시키는 데 사용하는 데이터가 인간이나 사회 문화가 가지는 편견과 오류를 그대로 포함한 상태를 말한다.

2 데이터 편향성 해소 방안

- 기계학습 모델을 학습시키는 데 사용하는 데이터셋을 구축하는 단계에서부터 편향성을 제거하는 것이 필요하다.
- 예측하고자 하는 대상의 다양성을 반영하고 있는지, 대상별 데이터의 양은 유사한지, 왜곡된 데이터가 없는지 등을 비판적으로 살펴보며 데이터셋을 구성하는 것이 필요하다.



신뢰할 수 있는 공정한 인공지능을 만들기 위해서는 어떻게 해야 할까요?



“새로운 기술의 영향력과 올바른 설계 방법에 대해 미리 고민하는 과정이 필요하다.”

2 윤리적 딜레마

Q 개념 탐색 주어진 문제 상황에서 어떤 선택을 해야 하는 걸까?

인공지능은 개발 목적에 따라 다양한 결과와 영향력을 가지고 올 수 있다. 예를 들어, 자율주행자동차의 경우에는 다양한 딜레마 상황에 놓일 수 있기 때문에 많은 사람들의 이해관계를 적절히 반영하는 것이 매우 어려울 수 있다.

인공지능이 가져올 수 있는 딜레마를 이해하고, 윤리적 인공지능 설계를 위해 고려해야 할 방안을 탐색한다.

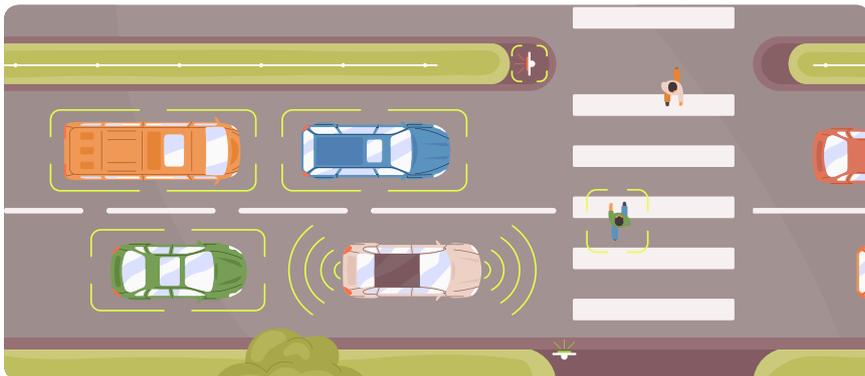
1 딜레마

주어진 문제 상황에서 선택해야 사항이 두 가지 중 하나로 정해져 있을 때, 그 어느 쪽을 선택해도 바람직하지 못한 결과가 나오게 되는 곤란한 상황을 가리킨다.

2 인공지능 딜레마

인공지능 시스템은 인간이 경험할 수 있는 딜레마와 동일한 상황 속에 노출될 수 있다. 하지만 인공지능 시스템을 어떻게 작동시키고 설계할 것인가는 결국 인간의 몫이기 때문에 딜레마에 대한 이해와 함께 지속적인 논의와 합의가 필요하다.

인공지능 딜레마!
여러분의 선택은
무엇인가요?



활동 미리 보기

핵심 개념과 관련된 활동은 어떤 내용으로 전개될까요?

활동 1

편향성은 어떻게
생기는 걸까?

- 데이터 편향성이 미치는 영향력은?
- 데이터 편향성을 해소하기 위한 방안은 무엇일까?



활동 2

인공지능의 선택은?
당신의 선택은?

- 인공지능 딜레마는 무엇일까?
- 나와 친구, 그리고 전 세계인들의 선택은 어떻게 다를까?
- 데이터 편향성과 딜레마에 대한 이해는 더 좋은 인공지능 세상을 만들기 위해 정말 필요할까?



2단계 창의융합적 사고로 해결하기

활동 1 편향성은 어떻게 생기는 걸까?

- ▶ 인공지능의 데이터 편향성이 수행 결과에 미치는 영향력을 이해하고, 데이터 편향성을 해소하기 위한 방안을 제시할 수 있다.

1 문제 이해하기

두 사람의 대화를 읽고, 그 문제가 왜 발생하는지 생각해 보자.



이번 활동에서 해결해야 할 문제는 무엇인지 써 보자.

.....

.....

무엇을 알아야 할까?

배경 지식

데이터 편향성

1. 데이터 편향성

편향성이란 통계적 예측 결과가 한쪽으로 치우치는 경향을 보임으로써 발생하는 오차를 말한다. 일반적으로 인공지능 알고리즘에 의한 기계적 예측이나 의사결정은 인간의 의사결정보다 편견이나 오류에서 자유롭다고 생각할 수 있다. 하지만, 인간에 의해 만들어지는 인공지능은 인간이나 사회 문화가 가지는 편견과 오류를 그대로 포함하고 있을 가능성이 있다.

2. 데이터 편향성 사례

한 연구에 의하면 인공지능 기반의 이미지 인식 시스템을 만들기 위해 인터넷에서 제공하는 데이터셋을 살펴보았더니 요리와 관련된 데이터의 경우에는 여성 데이터가 33% 이상 많다. 따라서 이를 활용한 인공지능 또한 대부분의 이미지를 여성으로 잘못 예측할 확률이 높아지는 것을 확인하였다.



COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	FRUIT



COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	MEAT



COOKING	
ROLE	VALUE
AGENT	WOMAN
FOOD	∅

이처럼 법률, 의료, 교육 서비스 등 다양한 곳에 활용되는 인공지능이 편향된 데이터셋만으로 작동하여 성별과 피부색 등에 따라 왜곡되게 예측한다면 차별 행위를 더욱 공고히 하게 되며 결국 인공지능 또한 신뢰를 잃을 수 있다.

3. 데이터 편향성에 대처하는 방안

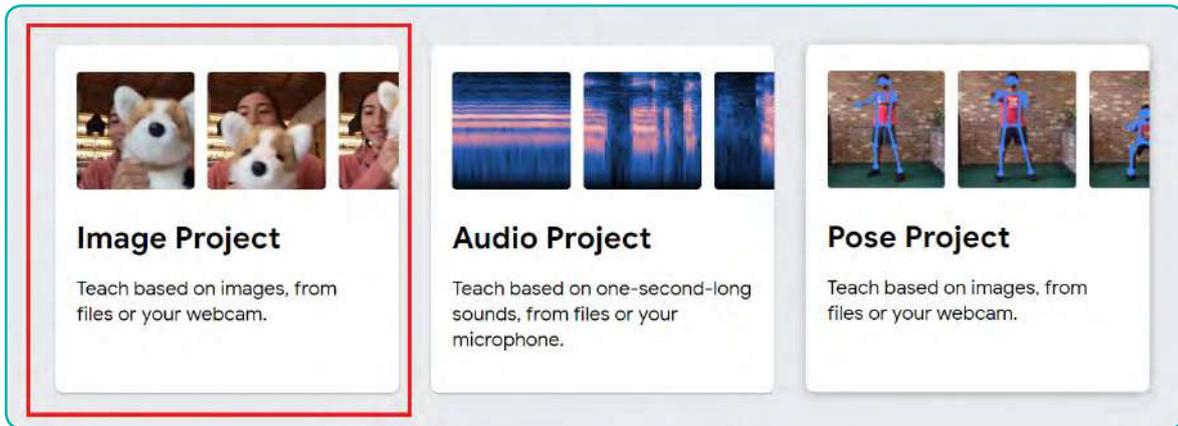
신뢰할 수 있는 공정한 인공지능을 만들기 위해서는 어떻게 해야 할까? 인공지능은 기계학습 모델을 설계하는 데 사용한 데이터에 의해 성능이 결정되는 경우가 많다. 따라서 예측하려는 대상의 다양성을 반영하고 있는지, 대상별 데이터의 양은 비슷한지, 왜곡된 데이터는 없는지 등을 비판적으로 살펴보고 데이터셋을 구성해야 한다.

2 문제 상황 체험하기

Teachable Machine을 활용해 인공지능이 어떠한 방식으로 편향성을 학습하게 되는지 알아보자.

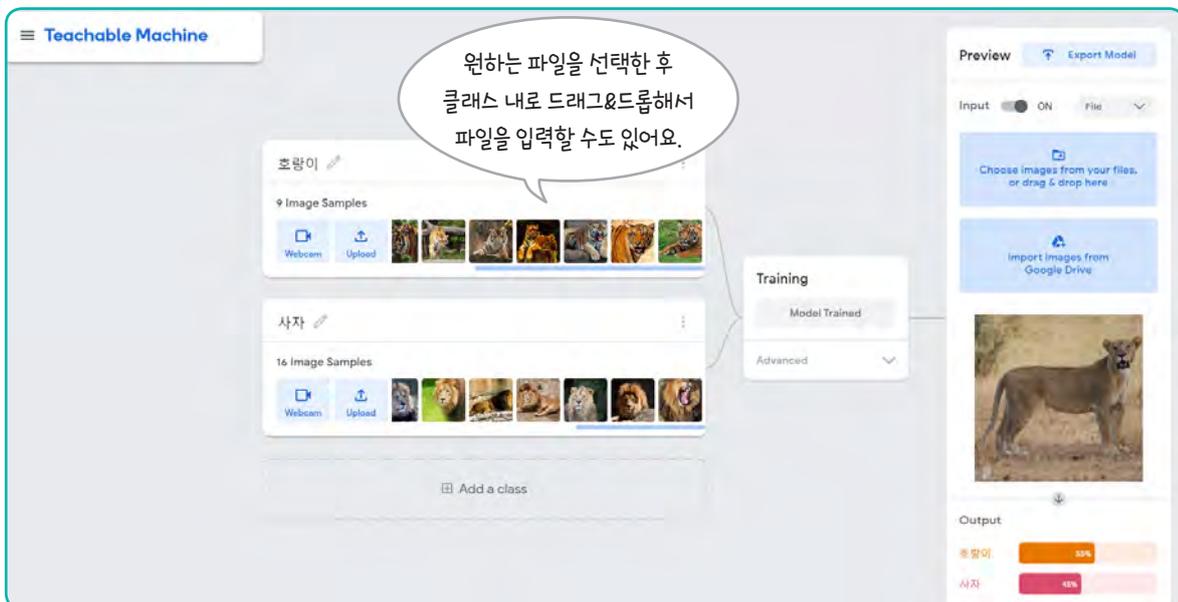
1 Teachable Machine 접속하기

<https://teachablemachine.withgoogle.com>에 접속하여 'Get Started'를 누른다. 데이터 형태를 선택하는 창에서 'Image Project'를 선택한다.



2 데이터 입력하기

- 1 소프트웨어 중심사회(www.software.kr)에 접속해 인공지능 윤리 체험 데이터를 다운로드한다. 데이터는 훈련 데이터, 테스트 데이터, 재구성 데이터로 구성되어 있다.
- 2 Teachable machine으로 이동해 Class1을 '호랑이', Class2를 '사자'로 작성한다. 티처블 머신에서는 대상별 데이터셋을 클래스(class)라고 부른다.
- 3 호랑이 Class에는 'upload'를 누른 후, 'choose images from your files'를 눌러 '훈련 데이터-호랑이' 폴더 내 9개의 이미지 파일을 선택하여 입력한다.
- 4 사자 Class도 3과 동일한 방법으로 '훈련 데이터-사자' 폴더 내 이미지 파일을 입력한다.



3 훈련시키기 & 테스트하기

- ① 데이터 입력하기를 마친 후, Teachable Machine의 ‘Training’ 창의 ‘Train Model’ 버튼을 눌러 훈련시킨다.
- ② 다운로드한 테스트 데이터를 Preview Input 영역에 입력해 Output에 나타나는 예측 성능을 확인하고 아래 표에 정리해 보자.

테스트 사진	Teachable Machine이 무엇으로 예측하였나?(호랑이? 사자?)	올바르게 맞힌 예측률은 얼마인가?	예측이 성공했나? (예, 아니요)
			
			
			
			
			
			

 **도움 자료**

훈련 데이터와 테스트 데이터 사용 시 주의 사항!

- 훈련 데이터와 테스트 데이터는 각각 용도가 다르다.
- 훈련 데이터는 이미지 분류 모델을 학습시킬 때 사용하는 데이터이다.
- 테스트 데이터는 학습을 마친 이미지 분류 모델이 잘 작동하는지 확인하는 데 사용하는 데이터이다.

3 문제 분석하기

Teachable Machine을 활용한 테스트 데이터 예측률을 확인하고 아래의 물음에 답해 보자.

1 예측률을 확인하고, 이와 같은 결과가 나온 이유를 아래 표에 써 보자.

테스트 사진	예측률은 얼마인가? 그 이유는 무엇일까?	테스트 사진	예측률은 얼마인가? 그 이유는 무엇일까?
	<ul style="list-style-type: none"> • 예측률: • 이유: 		<ul style="list-style-type: none"> • 예측률: • 이유:
	<ul style="list-style-type: none"> • 예측률: • 이유: 		<ul style="list-style-type: none"> • 예측률: • 이유:
	<ul style="list-style-type: none"> • 예측률: • 이유: 		<ul style="list-style-type: none"> • 예측률: • 이유:

• 훈련 데이터의 각 클래스별 이미지는 몇 개씩 구성되어 있을까? 또 어떤 문제가 있을까?

• 훈련 데이터의 각 클래스별 이미지는 얼마나 다양하게 구성되어 있을까? 어떤 문제가 있을까?

2 위 활동을 통해 인공지능이 가져올 문제를 예측해 보자.

3 다음 글과 영상 속에서 발견되는 문제는 무엇인지 생각해 보자.

인공지능은 우리 사회를 이롭게 만들 수 있는 무한한 잠재력을 가지고 있다. 이를 바탕으로 윤리적이고 공정한 인공지능을 개발해야 한다. 오늘날의 인공지능은 방대한 양의 데이터를 학습하는 기계학습을 기반으로 작동한다. 이때, 다양한 인공지능 시스템이 편향성을 가질 때 어떠한 문제를 가지고 올 수 있는지 알아보고, 이러한 문제를 어떻게 예방할 수 있는지 생각해 보는 것이 필요하다. 아래 영상은 인공지능이 우리 삶의 다양한 곳에 미치는 영향력을 살펴보고, 편향된 인공지능 설계를 예방하기 위해 필요한 윤리적 설계의 중요성을 안내한다.



[출처: <https://youtu.be/zNw5gJtHLc>]

• 영상에서는 제기하는 문제는 어떠한 것들이 있는가?

• 영상에서는 편향성을 개선하기 위해 어떠한 노력이 필요하다고 제안하는가?

4 더 나은 인공지능 만들어 보기

모둠별로 더 나은 인공지능 만들기를 한 후, 발표하며 결과를 공유해 보자.

1 더 나은 인공지능을 만들기 위한 나의 생각을 써 보자.

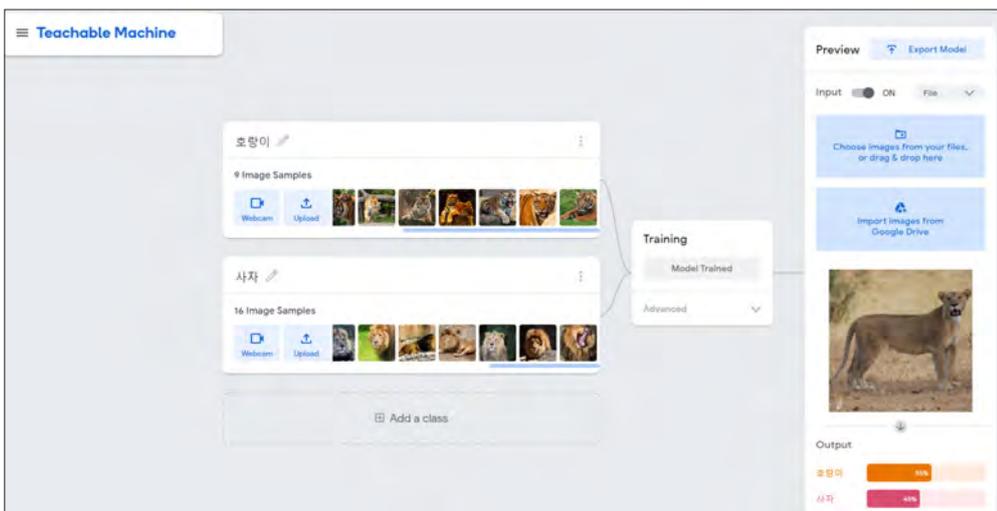
- 훈련 데이터를 구성할 때 어떤 조건들을 고려해야 할까?

예측하려는 대상의 다양성을 반영하기 위해 어떠한 고민들이 필요할지 구체적으로 생각해 봅시다!



2 모둠별 토론을 통해 더 나은 인공지능을 만들기 위한 조건을 작성해 보자.

3 모둠별 토론을 통해 나온 조건을 참고하여 훈련 데이터를 새롭게 수집하고 Teachable Machine으로 훈련시킨 후, 다시 한 번 테스트 데이터로 예측해 보자.



도움 자료

데이터 검색하기

- 인공지능의 성능을 개선하기 위해서는 '인공지능 윤리 체험 데이터'의 재구성 데이터를 활용할 수 있다.
- 그 외 더 필요한 데이터는 구글, 네이버, 다음 등의 검색엔진에서 이미지 검색을 활용해 데이터를 수집할 수 있다.

4 아래 표에 결과를 작성하여 이전보다 성능이 좋아졌는지 확인해 보자.

테스트 사진	예측은 성공했나?	예측률은 얼마인가?	기존보다 성능이 좋아졌나?
			
			
			
			
			
			

5 모둠별 발표를 통해 서로의 훈련 데이터의 구성 조건과 개선된 성능을 공유해 보자.
그리고 내가 속한 모듬의 훈련 데이터 구성 조건을 보완해 보자.

5 정리하기

1 데이터 편향성 체험과 개선 활동을 통해 새롭게 알게 된 점을 써 보자.



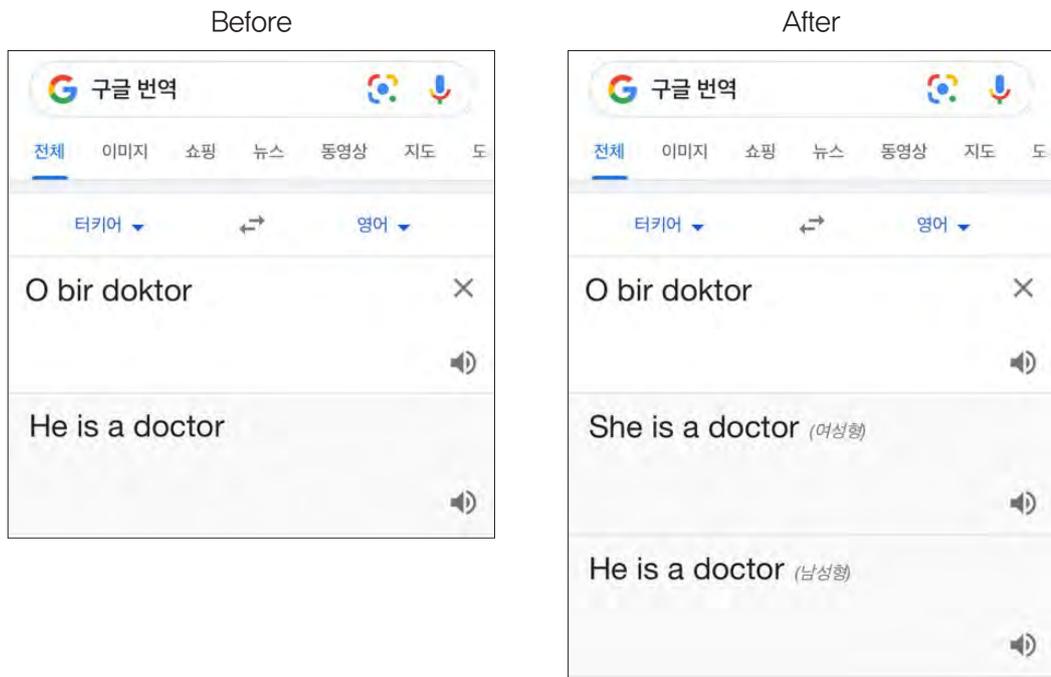
2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
편향성의 개념	편향성이 무엇인지 설명할 수 있는가?					
편향성 발생 원리	편향성이 어떻게 발생하는지 설명할 수 있는가?					
편향성 개선	편향성을 개선하기 위한 방안을 설명할 수 있는가?					

구글은 2019년 6월 ‘구글 AI 원칙’을 발표했다. 이는 AI 연구와 개발 단계에서 권고되는 원칙으로, 구글이 출시하는 AI 서비스는 모두 여기에 맞춰 평가한다. 구글은 인간이 중심이 되는 AI를 중시하며 인명 피해와 인권 침해가 발생하지 않는 범주로 사용 범위를 한정했다.

구글 AI 원칙에는 ‘불공정한 편견을 만들거나 강화하지 않는다.’에 기계학습 공정성이 담겨 있다. 순다 피차이(Sundar Pichai) 구글 최고 경영자(CEO)는 “인종, 민족, 성별, 국적, 소득 수준, 성적 지향, 장애, 정치적·종교적 신념과 관련된 부당한 영향이 미치지 않도록 노력해야 한다.”라고 말했다.

이러한 신념은 구글 번역 웹사이트에도 나타나고 있다. 성 중립적인 단어를 번역하면 여성형과 남성형 두 가지를 모두 보여 주는 방식으로 개선되었다. 이전에는 AI가 알아서 성별을 나타내는 대명사를 판단하여, 의사는 남성으로, 간호사는 여성으로 지칭하는 경우가 있었다.



이 밖에도 개방형 이미지 데이터 셋을 공유하는 등 공정성을 높이는 다양한 오픈소스 툴을 개발하고 있다. 구글 AI 리서치 프로그램 매니저 또한 “구글 제품은 많은 인구가 사용하며 모두가 접근할 수 있는 플랫폼을 지향하기 때문에 기계학습에 있어 공정성은 중요한 문제이다.”라고 말했다.

[출처: 인더뉴스, 2019. 6. 25. <http://inthenews.co.kr/article-3674/> 발췌]

활동 2

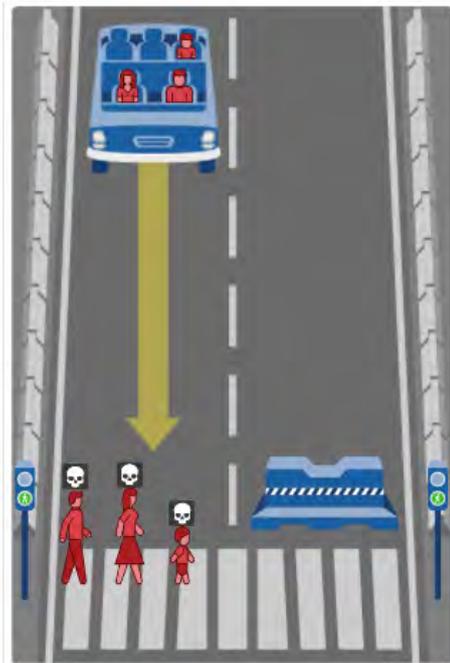
인공지능의 선택은? 당신의 선택은?

- ▶ 인공지능의 윤리적 딜레마를 이해하고, 인공지능을 설계하는 데 있어 충분한 사회적 논의와 합의가 필요함을 설명할 수 있다.

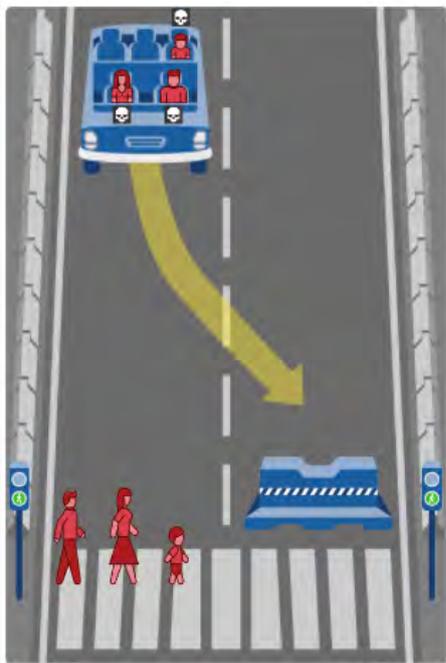
1 문제 이해하기

다음 두 개의 상황을 보고, 여러분은 어떤 선택을 할지 생각해 보자.

3명이 탑승하고 있는 자율주행자동차가 갑자기 나타난 장애물을 발견했을 때, 선택할 수 있는 상황은 아래와 같다. 자신의 도덕적 가치에 근거하여 한 가지 상황을 선택해 보자.



자율주행자동차 탑승객인 3명(어른 2명, 아이 1명)의 목숨을 구한다(단, 보행자인 어른 2명과 아이 1명은 사망한다).



보행자인 3명(어른 2명, 아이 1명)의 목숨을 구한다(단, 자율주행자동차 탑승객인 어른 2명, 아이 1명은 사망한다).

자신의 선택과 그 이유를 써 보자.

- 나의 선택은? _____
- 그 이유는? _____
- _____

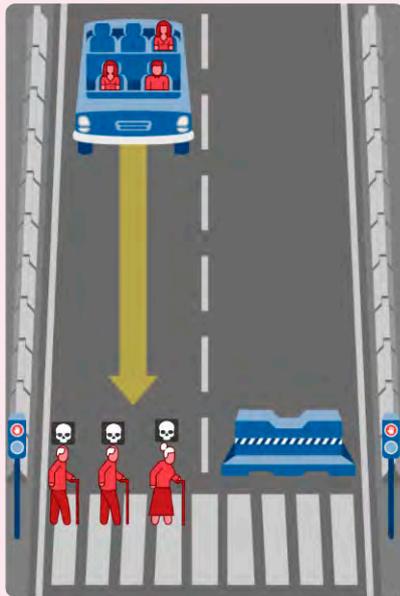
🎯 무엇을 알아야 할까?

🔍 배경 지식

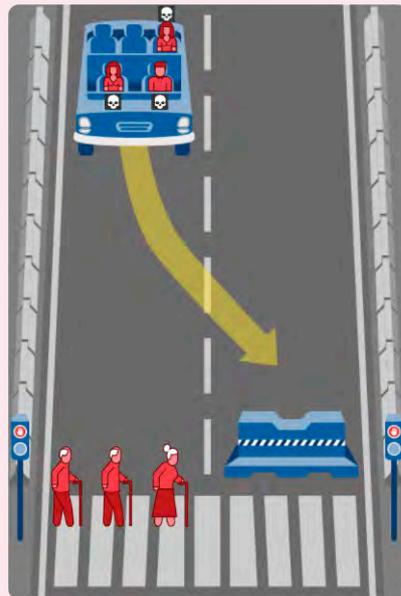
인공지능의 윤리적 딜레마

인공지능이 우리 삶 속에 들어오면서 의사결정을 하게 되는 경우가 많이 발생한다. 하지만, 인공지능이 내린 판단이나 행위가 항상 명확한 정답을 가지고 있거나 가치중립적이라고 보기 어려운 경우가 많다.

대표적인 사례로 자율주행자동차를 들 수 있다. 자율주행자동차는 운전자에게 편리함을 제공하고 교통사고를 줄이기 위해 다양한 정책을 반영하여 개발되고 있으나, 운전 중 무수히 많은 의사결정 상황에 놓이게 된다. 특히, 보행자가 갑작스럽게 등장하는 상황에서 운전자와 보행자 중에서 누구를 희생시킬지 또는 누구를 보호해야 할지 선택해야 하는 경우가 발생할 수 있다.



왼쪽 선택: 보행자인 노인 남성 2명, 노인 여성 1명이 사망할 수 있는 상황



오른쪽 선택: 탑승객인 남성 1명, 여성 2명이 사망할 수 있는 상황

자율주행자동차가 이러한 도덕적 결정을 내리는 데 기초가 되는 알고리즘을 정의하고 개발하는 것은 결국 인간의 과제이다. 특히, 자율주행자동차의 알고리즘은 지리, 문화, 경제적 요소 등에 따라 매우 다르게 개발될 가능성이 있다. 즉, 자율주행자동차를 만드는 데 한가지의 윤리 규칙을 적용하는 것은 힘들 수 있으며, 다양한 사회 문화를 고려한 적극적이고 지속적인 논의와 합의를 통한 알고리즘 설계가 필요하다.

2 문제 상황 체험하고 기록하기

모럴 머신(<https://www.moralmachine.net/hl/kr>)에 접속해 상단의 평가(Judge)를 클릭한 후 13가지 시나리오를 체험해 보자.



1 위의 그림에서 등장인물과 번호를 확인하고, 13개 시나리오를 간단히 요약한 후 자신의 판단 결과를 기록해 보자.

번호	시나리오 요약(희생자)		나의 선택과 이유
	직진할 경우	왼쪽으로 꺾을 경우	
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

2 13개 시나리오를 체험한 후 모럴 머신에 제시된 자신의 판단 결과를 기록해 보자.

판단	모럴 머신 결과
가장 많이 살려 준 캐릭터	
가장 많이 희생된 캐릭터	

3 토론 및 공유하기

자신의 시나리오 선택과 판단 결과를 친구들과 비교해 보자.

1 모둠원이 가장 많이 살려 준 캐릭터와 가장 많이 희생된 캐릭터를 적고, 그 이유를 써 보자.

가장 많이 살려 준 캐릭터는 _____ 이고,
가장 많이 희생된 캐릭터는 _____ 이다.

그 이유는 _____ 이다.
_____ 이다.

모럴 머신에 제시되는 시나리오 순서나 내용은 모두 다를 수 있어요. 시나리오 요약물 반드시 읽고, 같은 경우에는 서로의 판단 결과를 비교해 보세요.



2 13개 시나리오에서 가장 고민되었던 시나리오를 친구들과 공유하며, 친구들의 생각을 알아보자.

공유할 시나리오:

친구들 생각:

3 친구가 공유한 시나리오 중 찬성하는 시나리오와 반대하는 시나리오 1개씩을 제시하고, 그 이유를 간단히 서술해 보자.

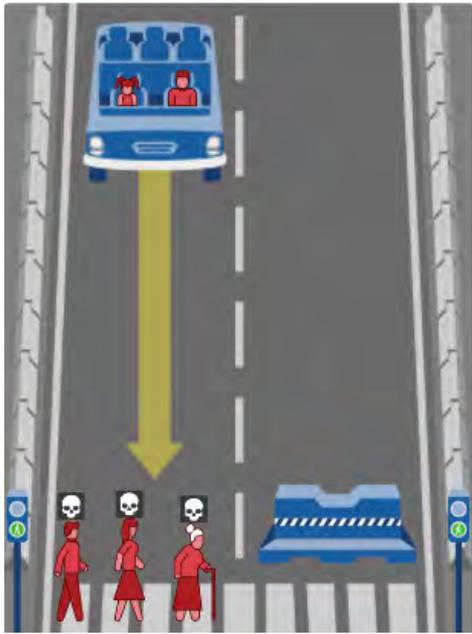
친구가 공유한 시나리오:

나의 생각:

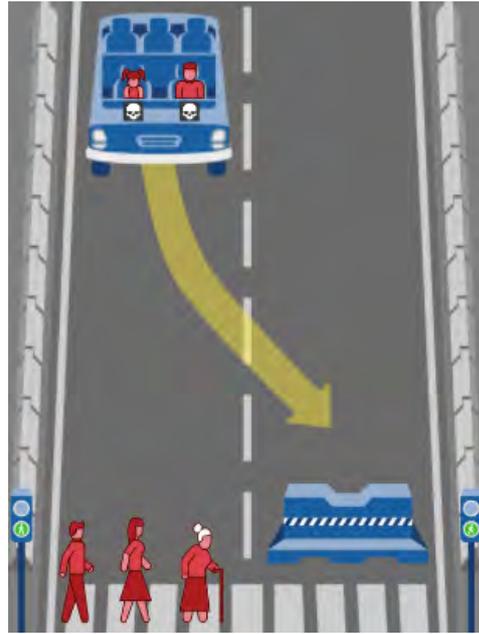
각자의 의사결정을 정리한 다음에는 친구들과 무엇이 같고 무엇이 다른지 이야기할 예정이니 신중히 작성해 주세요.



4 다음 시나리오를 보고, 친구들과 역할을 정해 보자.



자율주행자동차 탑승객인 어른 1명, 아이 1명의 목숨을 구한다(단, 보행자인 어른 2명과 노인 1명은 사망한다.).



보행자 3명(어른 2명, 노인 1명)의 목숨을 구한다(단, 탑승객인 어른 1명과 아이 1명은 사망한다.).

모둠 친구들과 개발사 CEO와 일반 시민 역할을 정한 후, 각 질문에 맞춰 이야기 나누어 보자.

- ① 자율주행자동차 개발 회사 CEO라면 위 그림을 보고 어떤 선택을 하는 자동차를 만들 것인가? 그 이유는 무엇인가?

• 결정과 생각은? 왜 그런 선택을 했나?

CEO는 한 명이 말아도 되고, 2명 이상이 함께 의사결정을 해도 됩니다.



- ② 일반 시민 입장에서 위 CEO가 생산하는 자동차를 구매할 의사가 있는가? 그 이유는 무엇인가?

• 결정과 생각은? 왜 그런 선택을 했나?

5 모럴 머신 체험을 통해 새롭게 알게 된 점을 써 보자.



도움 자료

MIT의 모럴 머신(Moral Machine) 실험 이야기

미국 매사추세츠공과대학(MIT) 연구진은 모럴 머신을 만들어 자율주행자동차의 딜레마 문제에 대해 233개 국가에서 수백만 명의 의견을 조사하였다.(2018년 10월 세계적인 과학저널 '네이처' 발표)

① 의견을 분석한 결과

세 가지 주요 공통 요소를 발견하였다.

- 첫 번째, 동물보다는 사람의 생명을 구하는 것을 우선으로 한다.
- 두 번째, 소수의 사람보다 많은 사람을 구하는 것을 우선으로 한다.
- 세 번째, 나이든 사람보다 젊은 사람의 생명을 구하는 것을 우선으로 한다.

② 세부적인 조사 결과

남성에 비해 여성을 구하려고 하였으며, 높은 지위에 있는 사람의 생명을 노숙자, 범죄자보다 조금 더 가치 있게 여기는 것으로 나타났다.

③ 동일한 사람을 대상으로도 규정(칙)을 지키는 보행자의 생명을 우선시 하는 것으로 나타났다.

④ 지역(나라), 성별, 연령 등에 따라 여러 딜레마에 대한 대응 결과가 매우 다르게 나타나는 양상 또한 확인하였다.

4 윤리적 설계 가이드 만들기

우리 모뎀만의 자율주행자동차 윤리적 설계 가이드를 만들어 보자.

- 1 다음 대화를 읽고, 딜레마 상황에 대처하기 위해 개인적인 특성(연령, 성별, 신체적 또는 정신적 체질)을 기준으로 한 알고리즘의 설계는 과연 적절한지 서로의 생각을 써 보자.

딜레마 상황을 대처하기 위해 최대한 명확한 기준을 세워서 자동차 설계를 해야 하지 않을까?



그렇지만, 나이나 성별로 기준을 세워버리면 그게 바로 차별이라고 할 수 있지 않을까?

자율주행자동차 윤리적 설계 가이드란 자율주행자동차를 개발하거나 사용할때 참고해야 할 규칙 또는 원칙을 말합니다.



- 2 다음 대화를 읽고, 다수보다 소수를 희생하는 알고리즘의 설계는 과연 적절한지 서로의 생각을 써 보자.

딜레마 상황에서는 무엇보다 다수를 살릴 수 있도록 설계하는 게 맞지 않을까?



하지만, 불법으로 도로를 건너는 두 명 때문에 안전 운전하는 사람을 희생시키는 건 더 큰 문제가 되지 않을까?

딜레마 상황에 대처하기 위해 정해져 있는 정답은 없습니다. 다른 사람의 생각을 존중하며 서로 이야기해 보세요.



3 지금까지의 내용을 참고해 우리 모둠만의 자율주행자동차 윤리적 설계 가이드를 만들어 보자.

제목: _____

• 핵심 원칙

예 첫째, 자율주행자동차는 성별, 나이, 인종, 장애 정도 등을 이유로 차별화해서는 안 된다.

• 설계자/개발자의 의무

예 자율주행자동차 설계자는 운행 중 사고 발생 시 책임을 명확히 하기 위해 사고 기록을 녹화할 수 있는 기능을 탑재해야 한다.

• 사용자의 의무

예 자율주행자동차의 안전 문제를 일으킬 수 있기 때문에 임의로 수정하거나 변경을 해서는 안 된다.

주어진 예시 외 최소 3가지 원칙을 만들어 봅시다.



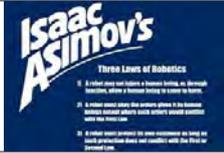
다른 모둠의 발표 내용을 조용히 경청하면서 다른 모둠의 아이디어를 메모해 보세요.



아이작 아시모프의 로봇 3원칙

로봇은 인간에게 해를 끼치지 않아야 하고 인간의 명령에 복종해야 하면서 로봇 자신의 존재를 보호하는 로봇 안전 준칙이다. 1942년 아이작 아시모프(Isaac Asimov)의 공상 과학 소설 '런어라운드(Runaround)'에서 처음 언급되었으며 세 가지 원칙은 다음과 같다.

- 첫째, 로봇은 인간에게 해를 가하거나, 또는 행동을 하지 않음으로써 인간에게 해를 끼치지 않는다.
- 둘째, 로봇은 첫 번째 원칙에 위배되지 않는 한 인간이 내리는 명령에 복종해야 한다.
- 셋째, 로봇은 첫 번째와 두 번째 원칙을 위배하지 않는 선에서 로봇 자신의 존재를 보호해야 한다.



1985년에 아시모프는 위 3대 원칙에 인류 집단 안전을 위해 0번째 법칙으로 '로봇은 인류에게 해를 가하거나, 해를 끼치는 행동을 하지 않음으로써 인류에게 해를 끼치지 않는다.'를 추가하였다.

국토교통부의 자율주행자동차 윤리 원칙

자율주행 윤리 가이드 라인(안) 개요

자율주행자동차의 도입에 따라 윤리적 판단을 기계가 대체하여 발생하는 결과에 따른 사회적 충격과 책임을 완화하기 위한, 판단의 목표와 가치를 제시한다.

주요 내용

○ 목표와 기본 가치(요약)

- 자율주행자동차의 목표는 인간의 안전과 복리 증진
- 인간이 안전하고 편리하며 자유로운 이동권 보장
- 인간의 생명을 동물이나 재산의 피해보다 우선적으로 고려(인간의 존엄성)
- 사고로 인한 개인적, 사회적 손실의 최소화(공공선)

○ 자율주행자동차와 관련된 각 주체들이 준수해야 하는 윤리 원칙(요약)

- 설계자의 의무: 자율주행자동차를 불법 개조하거나 임의로 시스템을 변경할 수 없도록 시스템 설계, 해킹을 방지할 수 있도록 자율 주행 자동차 설계 등
- 제작자의 의무: 제작·판매에 관련된 법규 준수, 자율주행자동차의 안전과 보안에 대한 보장 책임, 사용 연한 내의 유지보수와 결함에 대한 책임 등
- 관리자의 의무: 자율주행자동차 도입과 활용을 위한 사회적 인프라 확충 의무, 자율주행자동차 도입, 안전 및 모니터링 등에 관한 의무 등
- 소비자의 의무: 자율주행자동차 임의 개조·변경 금지, 오사용 및 불법적 사용으로 발생하는 문제에 대한 책임 의무, 법률 및 사용지침 준수 등



5 공유하고 정리하기

1 각 모둠에서 만든 윤리적 설계 가이드를 발표·공유하고 아래 표를 참고하여 최고의 가이드를 뽑아 보자.

※ 1점 ~ 5점으로 표시한다.

주요 가치	_____ 조				
인간의 안전과 복리를 증진시키기 위한 지침을 적절히 제공하였는가?					
사고로 인한 손실을 최소화시키기 위한 지침을 적절히 제공하였는가?					
다양한 이해관계자의 가치를 조화롭게 반영하였는가?					
총 합					

빈 공간에 모둠이 생각하는 가치를 써넣어 봅시다.



2 지금까지의 내용을 토대로 평가해 보자.

평가 요소	평가 항목	척도				
		1	2	3	4	5
딜레마 이해	자율주행자동차와 관련한 다양한 딜레마 상황을 설명할 수 있는가?					
자기 인식	자신이 생각하는 중요한 가치를 인식할 수 있는가?					
다양성 이해	자율주행자동차와 관련된 다양한 이해관계자의 입장을 설명할 수 있는가?					



인공지능 윤리 전 과정을 되돌아보며, 서로의 생각을 나누어 보자.

1 다음 글을 읽고, 아래 질문에 대해 생각해 보자.



"인간적인 인공지능을 만들기 위한 노력!"
인간적 기술 센터 (Center for Humane Technology)

인간적 기술 센터의 여정은 2013년 Google의 디자인 윤리학자인 트리스탄 해리스(Tristan Harris)가 유튜브와 인스타그램을 비롯한 유명 소셜 미디어의 비즈니스 모델이 막대한 부정적 영향을 가져올 수 있음을 관찰하면서 시작되었다.

그는 '사용자의 주의를 끌기 위한 요청(A Call to Minimize Distraction & Respect Users' Attention)'이라는 문서를 만들어 함께 일하는 개발자들에게 공유하며, 소셜 미디어가 사용자들의 주의를 지속적으로 끌기 위해 다양한 심리학과 인공지능 기술에 기반한 비즈니스 모델을 사용하고 있음을 지적하였다. 즉, 소셜 미디어가 사용자에게 필요한 정보를 제공하는 것에 목적을 두는 것이 아니라, 사용자가 소셜 미디어 서비스를 벗어나지 못하도록 지속적으로 자극적이고 편향된 정보를 제공하는 장치가 되고 있음을 폭로한 것이다.

이후, 2018년에는 소셜 미디어를 통해 선거가 조작될 수 있음을 폭로하며, 소셜 미디어에 숨어 있는 의도와 기술적 원리 등을 파악하여 활용할 것을 권장하고 있다. 또한 새로운 기술과 서비스를 설계할 때 여러 질문에 기초해 진행할 것을 권장하고 있다. 우리도 이러한 질문들을 한 번쯤 생각해 보며, 새로운 창작물들을 만들어 보자.

질문1. 사용자에게 맞춤형 정보를 제공할 때 나도 몰랐던 편견이 반영된 것은 아닐까?

질문2. 내가 만드는 기술이 우리 사회에서 발생하는 격차(불평등, 양극화)를 재생산하지 않을까?

질문3. 만약 내가 만든 기술이 글로벌 사용자를 고려한다면, 각 나라의 문화를 이해하거나 존중할 수 있을까?

질문4. 내가 만든 기술이 공동체 의식을 함양시키는 데 도움이 될 수 있을까?

2 추가 정보 조사하기

웹사이트를 통해 인간적 기술 센터에 대한 추가 정보를 찾아보자.

웹사이트: <https://www.humanetech.com/technologists>

3 나만의 질문 만들어 보기

인간적 기술 센터에서 제안하는 네 가지 질문 이외에 우리가 함께 고민해야 할 것들이 있을까? 나만의 질문을 만들어서 친구들과 공유해 보자.

친구들과 서로 내용을
공유해 보세요.



출처

I 지능 에이전트

- 대한민국 정부. (2019). 인공지능 국가전략.
- 블레클리 페인, 김한성 외 역. (2019). 모두를 위한 인공지능 윤리. KERIS.
- 스텐터 러셀, 피터 노빅. (2016). 인공지능: 현대적 접근방식(제3판). 제이펍.
- 유튜브. <https://www.youtube.com>
- A.I. Duet. <https://experiments.withgoogle.com/ai/ai-duet/view/>
- Auto Draw. <https://www.autodraw.com/>
- Experience Curiosity. <https://eyes.nasa.gov/curiosity/>
- The New York Times. Designed to Deceive: Do These People Look Real to You?. <https://www.nytimes.com/interactive/2020/11/21/science/artificial-intelligence-fake-people-faces.html>
- Research and Workshop report. (2019). Consumers International, Artificial Intelligence: consumer experiences n new technology.

II 인식

- 김환희. (2020). 텐서플로 2.0 프로그래밍. 위키북스.
- 동아일보. 로봇이 수저 놓고 '홈' 뎀 화상 코칭. <https://www.donga.com/news/It/article/all/20210116/104946309/1>.
- 사이트 고키. (2020). 밑바닥부터 시작하는 딥러닝2. 한빛 미디어.
- 삼성SDS. 스마트폰에는 어떤 센서가 탑재되어 있을까요? <https://www.samsungsds.com/kr/story/2020-sensor.html>.
- 아두이노. 아두이노 사이언스 저널. <https://www.arduino.cc/education/science-journal>
- 워드클라우드 생성기. <https://www.wordclouds.com>.
- 워드클라우드 생성기. wrdcloud.kr
- 이경님. (2017). 음성 언어 처리 기술, 어디까지 왔나, 국립국어원. https://www.korean.go.kr/nkview/nklife/2017_4/27_0405.pdf.
- 조경래. (2020). 처음 배우는 챗봇. 한빛미디어.
- 플리토. <https://ko.flitto.com/promotion/18>
- 허민석. (2020). 나의 첫 머신러닝/딥러닝. 위키북스.
- ITChosun. 딥러닝 기술로 '농산물자동 선별기 개발'. http://it.chosun.com/site/data/html_dir/2019/07/19/2019071901190.html
- LG CNS. 도시 안전 책임진다! 스마트시티 속 활약하는 자율주행 센서. <https://blog.lgcns.com/2400>
- TensorSpace. TensorSpace.js. <https://tensorspace.org>.
- ZDNetKorea. MS AI. 인간 수준의 이미지 묘사능력 구사. <https://zdnet.co.kr/view/?no=20201020152125>.

III 데이터

- 한국인터넷기업협회. (2018). 모바일 서비스 이용행태 조사 보고서
- 한국고용정보원. (2019). 4차 산업혁명 시대 내 직업 찾기
- 한국 아동·청소년 데이터 아카이브. <https://www.nypi.re.kr/archive/mps>

IV 기계학습

- 구글 인공지능 실험 사이트. <https://experiments.withgoogle.com/>
- 미 뉴욕주 학교에서 얼굴인식 기술 사용금지 (이미지 출처: 픽사베이) <https://zdnet.co.kr/view/?no=20201224082307>
- 인간과 인공지능의 토론. <https://blog.naver.com/innosociallab/221847033823>
- 피크비전. <https://www.peekvision.org/>
- Project Euphonia. <https://sites.google.com/view/project-euphonia/>
- 기계학습에서 사용하는 데이터. <https://medium.com/python-in-plain-english/write-a-simple-steganography-program-using-python-2e1a9fa785d4>

V 인공지능 윤리

- Center for Humane Technology. Together, we can change how technology is built. Retrieved from <https://www.humanetech.com/technologists>
- 로봇공학의 삼원칙. Retrieved from https://ko.wikipedia.org/wiki/%EB%A1%9C%EB%B4%87%EA%B3%B5%ED%95%99%EC%9D%98_%EC%82%BC%EC%9B%90%EC%B9%99
- AI+Ethics Curriculum for Middle School. Retrieved from <https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/>
- Moral Machine. Retrieved from <https://www.moralmachine.net/>
- Jieyu Z, et al. (2017). Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2979 - 2989. Association for Computational Linguistics.
- E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon, I. Rahwan (2018). The Moral Machine experiment. Nature. 563(7729)
- E. Awad, S. Dsouza, A. Shariff, I. Rahwan, J.-F. Bonnefon (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. Proceedings of the National Academy of Sciences, 117(5)

◎ 연구 · 집필진

- 김수환 총신대학교 교수
- 임진숙 북삼고등학교 교사
- 최정원 만월중학교 교사
- 장병철 선린인터넷고등학교 교사
- 김한성 소프트웨어정책연구소 선임연구원

◎ 총괄 · 기획

- 조성연 교육부 교육과정정책과장
- 장원영 교육부 교육연구관
- 김인주 교육부 교육연구사
- 이현숙 한국과학창의재단 팀장
- 성정숙 한국과학창의재단 연구원

◎ 편집 · 디자인

- 편 집 | (주)씨마스(권소민, 이은경, 최햇님)
- 디자인 | (주)씨마스(이기복, 광상엽)

고등학교 기초

학교에서 만나는
인공지능 수업

발행일 2021. 2. 10.
발행처 교육부 · 한국과학창의재단
편집 · 디자인 (주)씨마스
인쇄



고등학교 기초

학교에서 만나는
인공지능 수업



9 791163 693697
ISBN 979-11-6369-369-7