



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

텍스트 데이터 라벨링 기반
머신러닝 교육 프로그램 개발 및 적용

Development and application
of the Machine Learning educational
program based on text data labeling

제주대학교 교육대학원

초등컴퓨터교육전공

김 범 솔

2021년 8월

석사학위논문

텍스트 데이터 라벨링 기반
머신러닝 교육 프로그램 개발 및 적용

Development and application
of the Machine Learning educational
program based on text data labeling

제주대학교 교육대학원

초등컴퓨터교육전공

김 봄 솔

2021년 8월

텍스트 데이터 라벨링 기반
머신러닝 교육 프로그램 개발 및 적용

Development and application
of the Machine Learning educational
program based on text data labeling

지도교수 김 종 훈

이 논문을 교육학 석사학위 논문으로 제출함

제주대학교 교육대학원


초등컴퓨터교육전공


김 범 솔

2021년 5월

김봄솔의

교육학 석사학위 논문을 인준함

심사위원장 신 경 우 

심사위원 박 승 제 

심사위원 김 종 훈 

제주대학교 교육대학원

2021년 6월

목 차

국문 초록	i
I. 서론	1
1. 연구의 필요성	1
2. 연구 내용	2
3. 연구 절차	3
II. 이론적 배경	4
1. 인공지능(AI)	4
2. 머신러닝	5
3. 머신러닝 포 키즈	7
4. 데이터 라벨링	8
5. 컴퓨팅 사고력	9
6. 비버 챌린지	11
7. 계산적 인지력 검사	11
8. 선행연구 분석	12
III. 연구 내용	14
1. 요구 분석	14
2. 설문 내용	14
3. 교육 방향	17
4. 교육 내용	18
5. 교육자료 개발	20

목 차

IV. 연구 결과 및 해석	29
1. 연구 대상	29
2. 검사 도구	29
3. 연구 설계 및 처치	29
4. 사전-사후 검사의 정규성 검정	30
5. 컴퓨팅 사고력 사전-사후 검사 비교	31
6. 학생 산출물 심층 분석	32
7. 연구 결과 분석	38
V. 결론 및 제언	40
1. 결론	40
2. 제언	41
참고 문헌	43
ABSTRACT	46
부 록	48

표 목 차

〈표 I-1〉 ADDIE 모형에 따른 교육 과정 설계	3
〈표 II-1〉 문제해결을 위한 컴퓨팅 사고 단계	10
〈표 III-1〉 머신러닝에 대한 이해도	16
〈표 III-2〉 머신러닝 교육의 필요성	16
〈표 III-3〉 데이터 라벨링에 대한 이해	16
〈표 III-4〉 데이터 라벨링 교육의 필요성	17
〈표 III-5〉 교육 프로그램의 주제	17
〈표 III-6〉 교육 프로그램 강의 계획	19
〈표 III-7〉 언플러그드 활동 계획	20
〈표 III-8〉 텍스트 데이터 라벨링 실습 활동 계획	24
〈표 IV-1〉 실험 설계	30
〈표 IV-2〉 비버 챌린지 정규성 검정	30
〈표 IV-3〉 계산적 인지력 검사 정규성 검정	31
〈표 IV-4〉 비버 챌린지 대응 표본 t 검정 결과	31
〈표 IV-5〉 계산적 인지력 검사 대응 표본 t 검정 결과	32

그림 목 차

[그림 II-1] 인공지능, 머신러닝, 딥러닝의 관계	5
[그림 II-2] 지도학습, 비지도 학습, 강화학습의 정의	6
[그림 II-3] 머신러닝 포 키즈 프로젝트 예시	7
[그림 II-4] 인공지능 데이터 품질 표준안	9
[그림 III-1] 요구 분석 설문지 내용	15
[그림 III-2] 1단계 학습 과정 예시	21
[그림 III-3] 2단계 학습 과정 실습 결과물	22
[그림 III-4] 3단계 학습 과정 실습 결과물	22
[그림 III-5] 4단계 학습 과정 실습 결과물	23
[그림 III-6] 5단계 학습 과정 예시	23
[그림 III-7] 텍스트 데이터 라벨링 활동지	25
[그림 III-8] 텍스트 데이터 수집 예시	26
[그림 III-9] 텍스트 데이터 라벨링 예시	27
[그림 III-10] 데이터 입력 결과 예시	27
[그림 IV-1] 머신러닝 교육 프로그램 수업 장면	33
[그림 IV-2] 4~9차시 활동지 예시 1	34
[그림 IV-3] 4~9차시 활동지 예시 2	35
[그림 IV-4] 10~12차시 활동지 예시 1	36
[그림 IV-5] 10~12차시 활동지 예시 2	37

국 문 초 록

텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램 개발 및 적용

김 범 솔

제주대학교 교육대학원 초등컴퓨터교육전공
지도교수 김 종 훈

본 연구는 초등학생의 컴퓨팅 사고력을 향상하기 위한 교육 방법으로 텍스트 데이터를 활용한 데이터 라벨링 중심의 머신러닝 교육 프로그램을 개발하여 적용한 후 그 효과를 검증하였다. 교육 프로그램 개발을 위해 제주도 초등학교 교사 100명을 대상으로 사전 요구 분석을 진행하였고, 그 결과를 바탕으로 교육 설계 및 개발을 진행하였다. 개발한 교육 프로그램은 초등학교 6학년에 재학 중인 17명의 남녀 학생을 대상으로 6주간 12차시의 교육을 진행하였다. 해당 교육의 효과성을 검증하기 위해 '비버 챌린지(Bebras Challenge)'와 '계산적 인지력 검사(Computational Cognition Test) A, B형'을 활용하여 사전 사후 검사를 진행하여 교육적 효과를 분석하였다. 분석 결과 데이터 라벨링 중심의 머신러닝 교육이 초등학생의 컴퓨팅 사고력 향상에 이바지한 것으로 나타났다.

주요어 : 텍스트 데이터, 데이터 라벨링, 머신러닝 교육, 머신러닝 포 키즈, 컴퓨팅 사고력

I. 서 론

1. 연구의 필요성

4차 산업 혁명 시대가 가속화되면서, 4차 산업 혁명을 이끄는 IT 기술 교육의 중요성이 커지고 있다. 특히 인공지능(Artificial Intelligence: AI)과 관련된 기술의 발전은 실생활에 밀접한 여러 서비스에서 활용되며, 산업 구조에 걸친 패러다임에 영향을 주고 있다. 이런 시류에 맞춰 한국 정보과학교육연합회에서는 2020년 2월 6일, 초등학교 3학년부터 고등학교 2학년까지 단계별로 적용할 수 있는 '소프트웨어 교육 내용 개정안'을 발표하며, 최근 빠르게 성장하고 있는 인공지능 분야의 인재를 양성하기 위해 기존의 소프트웨어 영역에 '인공지능과 융합' 영역을 추가하였다고 밝혔다. 그리고 2019년 한국 창의 과학재단에서 발표한 차세대 소프트웨어 교육 표준모델에서는 인공지능 개념, 지식표현 및 추론, 머신러닝, 인공지능망에 관한 교육을 초등학교 3학년부터 고등학교 2학년을 대상으로 시행할 것을 제시하였다. 그러나 현재 적용되지 않고 있으며, 소프트웨어 교육과 융합하는 방향으로 '인공지능과 융합' 영역이 2019년에 도입되었다(한국 정보과학교육연합회, 2019).

또한 2020년 8월 과기정통부나 교육부 등 정부 관계부처에서 발표한 '전 국민 AI·SW 교육 확산 방안'을 살펴보면, 정보교육을 강화하여 초등학생부터 체계적으로 AI·SW 소양을 기를 수 있도록 1~4학년은 ICT 활용 교육을, 5~6학년은 AI·SW 교육을 체계적으로 실시하여, 4차 산업혁명에 대비한 필수역량 함양을 시키는 계획을 발표하였다. 또한 차기 교육 과정 개편 시 정보교육 수업 시간 확대를 추진을 목표로 하는 계획도 발표되었다(관계부처 합동, 2020).

그러나 현재 교육 현장에서는 2020년부터 적용된 2015년 개정 교육 과정을 중심으로 수업이 진행되고 있다. 정보교육의 경우, 초등학교 6학년이 실과 수업을 통해 배우는 소프트웨어 교육은 기존의 ICT를 활용한 소프트웨어의 기본 소

양을 익히는 대단원으로 구성된 교육 차시가 17차시로 배정되어 있다. 따라서 기존에 배정된 소프트웨어 교육 이외의 다른 영역을 다루기 어려운 실정이다. 그리고 담임교사가 소프트웨어에 대한 소양이 부족하여 양질의 소프트웨어 교육이 어려워 독립 교과 지정이 필요하다는 지적도 있다. 그리고 교육청 간 교육 수준 차이가 크게 발생하여, 이 격차를 줄이는 방안도 필요한 상황이다. 또한 인공지능과 소프트웨어 융합 능력에 대한 지도 교사들의 소양이 부족하여, 현장에서 실제 활용할 수 있는 교육 프로그램의 개발이 시급하다(한국 정보과학교육연합회, 2019).

본 연구에서는 이러한 교육 현장의 현실을 고려하여, 2015년 개정 교육 과정 내에서 초등학교 6학년을 대상으로 진행할 수 있는 총 12차시 분량의 머신러닝 교육 프로그램을 개발하였다. 교육 도구와 세부 주제는 사전 요구 분석 결과를 분석하여 선정하였다. 이와 같은 과정을 거쳐 개발한 교육 프로그램은 인공지능과 머신러닝의 개념을 전달하고, 나아가 머신러닝의 성능에 영향을 미치는 데이터 라벨링의 개념을 전달하여 학습자가 직접 데이터를 수집, 분류, 정제하는 과정을 거쳐 머신러닝 모델을 학습시키고, 학습시킨 모델을 활용하여 블록 코딩으로 인공지능 프로그램을 만들어보는 내용으로 구성하였다.

이렇게 개발한 교육 프로그램은 블록형 프로그래밍의 경험이 있으나 인공지능에 대한 개념을 학습한 적이 없는 초등학교 6학년 재학생을 대상으로 적용하였다. 그리고 교육 프로그램의 효과를 검증하기 위해 교육을 적용하기 전과 후에 비버 챌린지(Bebras Challenge)(2018)와 함께 김병수(2014)의 연구에서 개발한 계산적 인지력 검사 A, B형으로 사전·사후 검사를 진행하여 컴퓨팅 사고력의 향상 정도를 검증하였다.

2. 연구 내용

텍스트 데이터의 데이터 라벨링을 활용한 머신러닝 교육 프로그램 적용하기 위해서 다음과 같은 연구 과정을 거쳤다.

첫째, 머신러닝 교육 프로그램 중 초등학생을 대상으로 진행된 선행연구와 관련한 연구를 조사하고, 머신러닝 교육을 진행할 때 고려해야 할 사항을 정리하

여, 교육 프로그램의 주제를 선정하였다.

둘째, 현직 초등교사 100명을 대상으로 머신러닝 교육에 관한 사전 요구 분석 설문을 통해, 교육 현장에서 초등교사들이 지닌 머신러닝에 대한 인식과 머신러닝 교육을 할 때 고려되었으면 하는 도구나 내용에 대한 분석을 진행하였다. 이후 기존 연구 분석과 초등교사 대상으로 진행한 사전 요구 분석 사항을 바탕으로 초등학교 6학년 학생들에게 적합한 교육 프로그램 주제 및 활동을 구성하여 교육 프로그램을 개발하였다.

셋째, 위의 과정을 거쳐 개발한 교육 프로그램을 초등학교 6학년 대상으로 실제 적용을 진행하고, 교육 전과 교육 후의 사전·사후 검사를 통해 진행된 교육 프로그램의 교육적 효과를 검증하였다.

3. 연구 절차

본 연구에서는 ADDIE 모형의 절차에 따라 교수 체제 설계 과정의 일반적인 형태를 고려하여 교육 프로그램을 <표 I-1>과 같이 개발하였다.

<표 I-1> ADDIE 모형에 따른 교육 과정 설계

단계	프로그램 연구 및 개발 절차
분석 (Analysis)	· 요구 분석(초등교사 100명 대상 설문) - 머신러닝에 대한 인식 및 머신러닝 교육 프로그램 교수 도구 등 9개 문항
설계 (Design)	· 수행목표 명세화 - 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램의 효과 · 평가 도구 설계
개발 (Development)	· 교수자료 개발 - 12차시 교수학습 과정안, 교재 개발
실행 (Implementation)	· 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램 적용 - 교육 프로그램 개발 후 초등학생 6학년을 대상으로 사전 검사 후 교육 시행
평가 (Evaluation)	· 사후 검사 후 사전 검사와 비교하여 교육적 성과 점검 - 컴퓨팅 사고력과 계산적 사고력의 변화 정도 확인

Ⅱ. 이론적 배경

1. 인공지능(AI)

일반적인 컴퓨터 프로그램은 미리 프로그래밍이 된 코드에 따라 미리 입력된 결괏값을 보여준다. 예를 들어, 인터넷 쇼핑몰의 경우, 미리 입력한 상품 데이터와 해당 상품 데이터의 수량, 옵션 등을 미리 프로그래밍한 결과를 웹페이지를 통해 사용자에게 제공한다. 그에 반해 인공지능은 입력된 데이터의 규칙을 컴퓨터 스스로 찾아내거나 수정하는 학습 과정을 통해 추론한 결괏값을 보여준다. 최근 여러 인터넷 쇼핑몰에서 활용되고 있는 ‘추천 시스템’이 대표적인 인공지능을 활용한 서비스이다.

인공지능(Artificial Intelligence : AI)의 개념은 1956년에 개최된 다트머스 콘퍼런스에서 존 매카시(John McCarthy)가 창시했다. 존 매카시는 “the science and engineering of intelligent machines(지능이 있는 기계의 과학과 공상)”라고 인공지능을 정의했다. 이후 대화를 할 수 있는 컴퓨터인 엘리자(ELIZA)가 개발되었다. 엘리자의 여러 기능 중 치료사 역할을 맡은 대화 시뮬레이션 프로그램인 닥터(DOCTOR)가 가장 유명하다. 닥터는 환자 역할을 맡은 인간이 건네는 제시어에 따라 컴퓨터(엘리자)가 자동으로 응답해 주는 시스템이다(이영호, 2019).

이후 전문가 시스템(Expert System)이 등장하며 범용으로 활용되는 인공지능이 아니라 목적에 따라 인공지능을 개발하려는 시도가 시작되었다. 그러나 전문가 시스템은 “만약 X는 Y라면”이라는 규칙에 따라서 추론이 성립되도록 구성되어, 규칙에 없는 내용을 입력하면, 적절한 대응할 수 없는 한계점을 지니고 있다(오니시 가나코, 2019). 이러한 한계를 개선하고자 제시된 인공지능 프로그래밍 기술이 머신러닝이다.

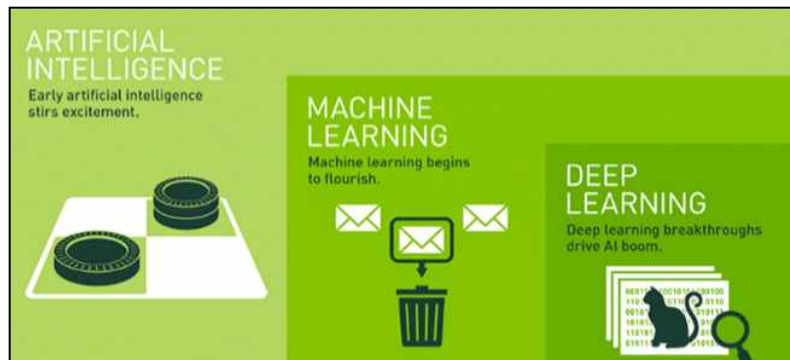
본 연구에서는 머신러닝의 개념을 전달하기 전, 인공지능이 탄생하게 된 배경

을 학습자에게 전달하고자 한다. 그러나 실제 교육 프로그램은 전문가 시스템을 주로 다루는 인공지능에 관한 내용보다 실제 추론을 기반으로 학습하는 머신러닝의 개념을 중점적으로 다룰 예정이다.

2. 머신러닝

머신러닝은 [그림 II-1]¹⁾과같이 인공지능에 속한 개념이다. 1959년 아서 사무엘(Arthur Samuel)은 머신러닝을 “기계가 일일이 코드로 명시하지 않은 동작을 데이터로부터 학습하여 실행할 수 있도록 하는 알고리즘을 개발하는 연구 분야”라고 정의하였다. 이러한 방식을 데이터 중심 접근 방식(Data-Driven Approach)이라고 한다. 이 방식으로 머신러닝은 인공지능의 전문가 시스템이 지닌 한계를 극복하여, 다양한 상황에서 유연하게 대응할 수 있었다(이영호, 2020).

현재 인공지능 서비스라고 소개되는 시스템은 머신러닝을 기반으로 개발된 방식이며, 인공지능으로 구현했다는 표현보다는 머신러닝으로 구현했다는 표현이 좀 더 정확한 표현이다(오니시 가나코, 2019). 그러나 일반 소비자들에게 익숙한 용어인 인공지능 또는 AI라는 표현을 활용하는 것이 대중에게 쉽게 와 닿아 인공지능으로 구현했다는 표현을 자주 사용한다.



[그림 II-1] 인공지능, 머신러닝, 딥러닝의 관계

1) NNIVIA KOREA. (2016). 인공지능과 머신러닝, 딥러닝의 차이점을 알아보자 (https://blogs.nvidia.co.kr/2016/08/03/difference_ai_learning_machinelearning/)

머신러닝은 컴퓨터가 학습하는 방식은 [그림 II-2]와 같이 지도학습, 비지도 학습, 강화학습으로 분류한다. 지도학습은 정답이 있는 데이터를 활용하여 학습을 진행하며, 훈련 데이터를 학습한 결과를 바탕으로 테스트 데이터의 정답을 맞추는 방식으로 학습 모델이 구성된다. 비지도 학습은 주어진 데이터를 비슷한 유형으로 나눠 군집화(Clustering)를 하며, 기계가 스스로 데이터의 특성을 파악하는 학습 방식이다(장준혁, 2018). 강화학습은 대중에게 인공지능을 알린 알파고 학습하는 방식으로, 특정 조건에서 성공했을 때 일정한 점수를 주어, 강화된 패턴을 익히도록 설계된 학습 방식이다.



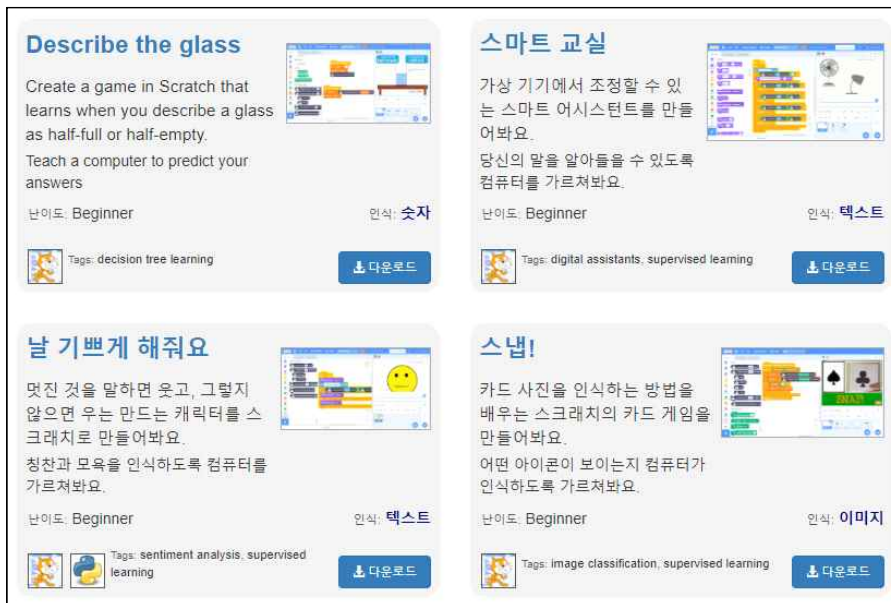
[그림 II-2] 지도학습, 비지도 학습, 강화학습의 정의

지도학습은 현재 활용하는 인공지능 서비스 중에서 가장 많이 활용하는 학습 방법으로 아마존에서 시작한 사용자별 추천상품 서비스를 필두로, 넷플릭스, 유튜브, 왓챠의 사용자별 콘텐츠 추천 서비스 등에서 활용하고 있다. 앞으로 진행될 교육 프로그램은 실제 생활에서 인공지능 서비스로서 많이 접할 수 있는 지도 학습을 기반으로 구성할 예정이다.

2) <https://jeongminhee99.tistory.com/68>(머신러닝의 이해와 지도학습을 이용한 분류(개념))

3. 머신러닝 포 키즈

‘머신러닝 포 키즈’는 IBM에서 제공하는 머신러닝 교육 서비스이다(Machine Learning For Kids, 2020). IBM에서 개발한 머신러닝 시스템인 왓슨(Watson)을 활용하여, 숫자, 텍스트, 이미지, 소리 데이터 중 한 가지 데이터 방식을 선택하여, 원하는 데이터를 입력하여 머신러닝을 체험할 수 있는 프로그램으로 구성되어 있다(이영호, 2020). 머신러닝으로 활용할 수 있는 모델은 ‘워크시트’ 메뉴에 공개된 프로젝트를 통해 제시되어 있다. 2020년 기준으로 총 37개의 프로젝트가 등록되어 있으며, 각 프로젝트는 인식하는 데이터의 종류, 난이도, 프로젝트에 사용하는 프로그래밍 언어에 따라 구분되어 있다. 머신러닝 포 키즈를 활용하는 선생님과 학생은 ‘워크시트’ 메뉴에서 [그림 II-3]³⁾과 같이 제시된 프로젝트 중 하나를 선택하여 실습을 진행할 수 있으며, 선택한 프로젝트의 모델을 응용하여 변형된 프로젝트를 만들 수도 있다.



[그림 II-3] 머신러닝 포 키즈 프로젝트 예시

3) <https://machinelearningforkids.co.uk/?lang=ko#!/worksheets> (머신러닝 포 키즈)

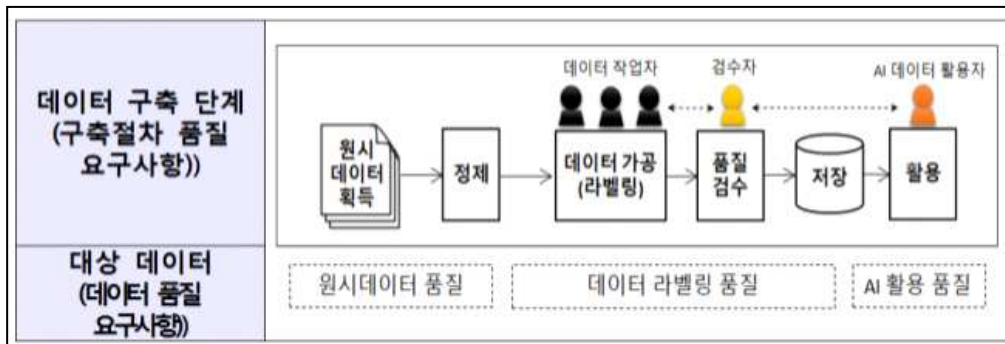
구체적인 활용 방법은 다음과 같다. '워크시트' 메뉴에서 실습을 원하는 프로젝트를 선택한 뒤, '프로젝트' 메뉴에서 해당 프로젝트의 이름을 넣고, 훈련하고자 하는 데이터 타입과 데이터에 사용된 언어를 선택하면, 모델 구성을 완료할 수 있다. 모델을 구성한 뒤 학습자가 데이터의 라벨을 직접 정하고, 학습할 데이터를 수집한다. 라벨링 한 데이터가 일정량 수집되면, 학습을 진행할 수 있다. 학습이 완료된 뒤에는 테스트 데이터를 넣어 모델의 정확도를 테스트할 수 있다.

본 연구에서는 머신러닝 포 키즈에서 제공하는 모델 중 텍스트 데이터를 활용한 모델을 선정하여, 텍스트 데이터의 수집, 분석 및 분류를 활동지를 통해 진행하고, 정제된 텍스트 데이터를 머신러닝 포 키즈의 프로젝트에 입력하여, 학습시키는 방식으로 교육을 진행할 예정이다.

4. 데이터 라벨링

데이터 라벨링은 머신러닝의 학습에 사용할 데이터 샘플을 검출하고 태그(라벨)를 붙이는 과정을 말한다. 실무에서는 전문 소프트웨어를 활용하여 라벨링을 하거나 인력으로 라벨링을 진행한다. 데이터 라벨링은 머신러닝의 학습을 위한 중요한 사전 작업이다. 지도학습의 경우 입출력 데이터에 모든 분류 라벨을 부착하여, 머신러닝 모델이 학습을 진행할 때 정답 여부의 근거를 제공한다 (Margaret R, 2019). 데이터 라벨링 과정에서 발생한 오류는 머신러닝의 정확도를 떨어뜨린다. 따라서 정확한 데이터 라벨링은 머신러닝의 성능을 높이는 데 기여하고, 이를 위해 데이터 수집 및 라벨링 과정을 이해하고 기준을 세우는 것이 중요하다.

2020년 10월에 과학기술정보통신부에서 발표한 인공지능(AI) 데이터 품질 표준안에서도 [그림 II-4]와 같이 데이터 라벨링 품질을 검수하는 과정을 제시하고 있다. 이는 머신러닝 모델의 성능에 영향을 미치는 데이터의 품질에 데이터 라벨링 과정이 지닌 중요성을 잘 보여주는 사례이다.



[그림 11-4] 인공지능 데이터 품질 표준안

이보배는 데이터 불균형이 머신러닝 모델 성능에 미치는 영향에서 데이터 라벨링 단계에서 불균형한 데이터가 제공되었을 때 머신러닝 모델 성능이 어떻게 달라지는지 연구하였다. 데이터 라벨링 단계에서 데이터의 불균형이 클수록 머신러닝 모델의 성능에 영향이 있음을 확인할 수 있었다(이보배, 2020).

이처럼 머신러닝을 위해 필요한 데이터를 수집하고 적절하게 분류하는 데이터 라벨링은 머신러닝의 성능에 영향을 미치는 중요한 과정으로, 초등교육과정에서 머신러닝 교육이 진행될 때도 비중 있게 다뤄져야 하는 영역이다.

5. 컴퓨팅 사고력

윙(Wing)이 주창한 컴퓨팅 사고력은 소프트웨어 교육 시 교수자가 학습 전략을 수립할 때 고려해야 하는 핵심 역량으로 평가받고 있다. 윙(Wing)은 컴퓨팅 사고력을 추상화와 분해를 통해 복잡한 시스템을 설계하거나 어려운 문제를 해결하는 것이라고 정의하고, 이러한 컴퓨팅 사고력의 핵심은 프로그래밍이 아닌 개념화에 있으며, 단순 반복적인 기술이 아닌 모든 사람이 갖춰야 하는 핵심 역량이라 기술한다.

ISTE(International Society for Education)의 Computational Thinking leadership toolkit에서는 컴퓨팅 사고력의 세부 요소로서 자료 수집, 자료 분석, 자료 표현, 문제 분해, 추상화, 알고리즘과 절차, 자동화, 시뮬레이션, 병렬화 등 9가지를 제시하였다.

우리나라도 소프트웨어 교육 시 컴퓨팅 사고력을 기르는 방향으로 교육 전략을 수립하고 나아가 기존의 교과목인 수학, 과학 등과 융합하여 컴퓨팅 사고력을 향상하기 위한 연구가 계속 진행되고 있다. 이영준은 초·중등단계에서 컴퓨팅 사고력 도입을 위한 기초 연구에서 초·중등 교육 과정에 CT를 도입하기 위해 교과별 CT 교육 프로그램을 개발하여 제시하였다(이영준, 2014). 교육부에서도 2015 소프트웨어 운영지침 해설서를 통해 <표 II-1>과 같이 문제해결을 위한 컴퓨팅 사고력 단계와 각 단계의 구성요소를 제시하였다(교육부, 2015).

<표 II-1> 문제해결을 위한 컴퓨팅 사고 단계(교육부, 2015a)

구성요소		정의
	자료 수집	문제해결에 필요한 자료 모으기
	자료 분석	자료의 이해, 패턴 찾기, 결론 도출하기
	구조화	문제를 그래프, 도표, 그림 등으로 시각화하기
추상화	분해	문제를 관리 가능한 수준의 작은 문제로 나누기
	모델링	문제해결을 위한 핵심 요소를 추출하고, 모델 만들기
	알고리즘	문제를 해결하기 위한 일련의 단계를 알고리즘으로 표현하기 (절차적 표현)
자동화	코딩	프로그래밍 언어를 이용해 문제해결 과정을 자동화하기
	시뮬레이션	프로그램(소프트웨어)을 실행하기
	일반화	문제해결 과정을 다른 문제에 적용하기

머신러닝 교육 또한 소프트웨어 교육의 한 분야로서 학습자의 컴퓨팅 사고력을 향상하는 방향으로 학습 전략을 수립하여 교육 프로그램을 개발하려는 다양한 연구가 진행 중이다. 본 연구에서는 머신러닝 교육 프로그램을 개발할 때, 컴퓨팅 사고력 단계와 각 구성요소를 고려하여, 학습자가 해당 교육 프로그램을 학습한 뒤 컴퓨팅 사고력이 향상될 수 있게 교육 프로그램을 구성할 예정이다.

6. 비버 챌린지

비버 챌린지(Bebras Challenge)는 학생들의 컴퓨팅 사고력을 평가하기 위해 개발된 과제 해결 중심의 평가 모델이다. 김은지는 '컴퓨팅 사고력 평가 도구로써 비버 챌린지 2017 신뢰도 분석 : 한국 그룹 III를 중심으로'에서 비버 챌린지 문항의 신뢰도가 Cronbach's Alpha = .784로써 지능검사 같은 인지적 검사에서 .7을 쓸 수 있다는 Kline의 주장을 소개하며, 문항의 신뢰도가 높다고 평가하였다(김은지, 2018). 이는 비버 챌린지가 컴퓨팅 사고력을 일관성 있게 측정할 수 있다는 것을 의미한다.

본 연구에서는 박은주의 알고리즘 교육을 통한 비버 챌린지 결과 분석에서 제시한 문항별 난이도를 반영하여(박은주, 2019), 비버 챌린지에서 제시하는 채점 기준에 맞춰 채점하였으며, 검사 도구로 비버 챌린지를 사용하여 사전-사후 검사 통제집단설계(pre test-post test control group design)를 바탕으로 연구를 진행하였다.

7. 계산적 인지력 검사

김병수는 계산적 사고력 신장을 위한 PPS기반 프로그래밍 교육 프로그램에서 계산적 사고의 조작적 정의를 Wing이 주창한 컴퓨팅 사고력의 정의를 활용하고 있음을 밝혔다. 또한 컴퓨팅 사고력과 관련된 선행연구를 바탕으로 계산적 사고력을 분석하여 제시하였다(김병수, 2014). 이와 같은 연구 내용을 살펴볼 때, 계산적 인지력 검사를 컴퓨팅 사고력을 측정하는 도구로 활용할 수 있음을 알 수 있다. 이에 본 연구에서는 비버 챌린지와 함께 계산적 인지력 검사 A, B형을 활용하여 컴퓨팅 사고력을 측정하였다.

김병수의 연구에서 개발한 계산적 인지력 검사(Computational Cognition Test) A, B형은 계산적 사고의 구성요인을 계산적 인지력(추상적 사고, 비판적 사고, 논리적 사고, 재귀적 사고, 알고리즘적 사고)과 계산적 창의성(창의적 사고)으로 구분하여 설정하여 개발된 검사 도구이다(김병수, 2014). 신뢰도 측정

은 문항 내적 일관성에 대해 측정하였고 Cronbach's α 값으로 신뢰도를 평가하였다.

신뢰도평가의 Cronbach's α 계수의 최소치를 .5로 정하고 이에 따라 각 문항의 신뢰도를 분류된 검사로서, A형 검사, B형 검사 간 계산적 인지력 지수(각 문항의 총점)의 상관계수가 .931로 매우 높은 상관을 지니고 있으며, 두 검사지 A형, B형은 동형 검사를 이용하여 신뢰도를 확보된 검사이다. 또한 계산적 인지력 검사 A형과 GALT 상관 분석의 상관계수가 .615이고, 검사 B형과 GALT의 상관계수는 .631로 논리적 사고력을 측정할 수 있다는 타당성을 확보하고 있다.

8. 선행연구 분석

김갑수는 초등학생의 인공지능 교육을 위한 교수학습 모델 개발 및 적용에서 다섯 단계로 인공지능 학습 모델을 제시하였다. 첫 번째는 문제 이해 단계로서 주어진 문제를 인공지능이 해결할 수 있는지 확인하는 단계이고, 두 번째는 데이터를 준비하는 단계로서 주어진 문제에 적절한 데이터를 선정하고 수집하는 단계이다. 세 번째는 인공지능 모델을 결정하고, 네 번째는 실제 인공지능 프로그램을 작성하는 단계로 자기조절 학습 모델의 4C(Copying, Changing, Creating, Challenging)을 활용한다. 마지막 5단계에서는 보고서를 작성하는 모델을 제안하였다. 파이썬을 활용하여 인공지능 프로그램 교육을 하였으나, 초등학생들이 파이썬을 이해해야 한다는 점과 교육을 위해 많은 시수가 필요하다는 점을 제언했다(김갑수, 2017).

본 연구에서는 김갑수의 연구에서 초등학생들이 파이썬을 먼저 이해하고 있어야 해당 인공지능 교육 프로그램을 익힐 수 있다는 점에 착안하여, 인공지능 교육 프로그램을 진행할 때, 초등학생들이 소프트웨어 교육 과정에서 배우는 블록 코딩을 활용할 수 있는 교육 프로그램인 머신러닝 포 키즈를 선택하였다.

이은경은 국내외 초·중등학교 인공지능 교육 과정 분석에서 인공지능 교육 과정 개발을 위한 방향 및 단계를 구체적으로 설정할 필요가 있음을 제안한다. 미국이나 EU의 경우 별도의 인공지능 교육을 위한 교육 과정이나 내용을 제시하고

있으나, 한국은 차세대 소프트웨어 교육 과정 표준모델에 인공지능 관련 영역을 포함함으로써 AI 교육을 위해 필수적인 교육 요소들을 포함하지 못하는 한계를 지니고 있다고 제언했다(이은경, 2020).

본 연구에서는 이은경의 제언을 바탕으로 본 교육 과정 방향과 단계를 좀 더 구체적으로 설정하기 위해, 머신러닝에서 다루는 다양한 데이터 형식 중 텍스트 데이터를 중심으로 교육을 구성하였다. 그리고 학습자가 스스로 데이터를 수집하고, 수집한 데이터의 라벨링을 기준을 세우고, 데이터 라벨링을 진행한 뒤, 모델을 학습시키고, 평가하는 과정을 기록할 수 있도록 구성하여, 학습자가 머신러닝 모델을 개발하는 과정을 단계적으로 익힐 수 있게 구성하였다.

또한 이영호는 블록형 프로그래밍 언어 기반 인공지능 교육이 학습자의 인공지능 기술 태도에 미치는 영향 분석에서 블록형 프로그래밍 언어 기반의 인공지능 교육이 학생들이 가지고 있는 인공지능 교육에 대한 태도에 정적인 방향으로 영향을 미친다고 밝히고 있으며, 실제 인공지능 기술을 체험해봄으로써 인공지능 기술에 대한 심리적 장벽을 낮출 수 있었다고 밝히고 있다(이영호, 2019).

본 연구에서는 이영호의 연구 결과를 바탕으로 학습자에게 인공지능 기술에 대한 심리적 장벽을 낮추기 위해, 교육 초기에 인공지능과 관련된 개념을 전달할 때, 실생활에서 접할 수 있는 머신러닝 서비스를 소개하면서, 머신러닝의 개념과 구현되는 원리를 전달하는 방식을 선택하여 구성하였다.

이처럼 본 연구에서는 제시한 선행연구들의 제언을 확장하는 방향으로 기존의 머신러닝 교육에서 다루지 않았던 데이터 라벨링에 중점을 두었다. 데이터를 준비하는 단계에서 주어진 문제에 적절한 데이터를 선정하고 수집하며, 머신러닝의 구성과 성능을 높이는 방법에 대해 학습자가 고민하고 수정할 수 있도록 교육 프로그램을 구성하였다. 개발한 교육 프로그램을 머신러닝 교육에 대한 경험이 없는 초등학교 6학년 대상으로 적용하여, 교육 시행 후 컴퓨팅 사고력에 미치는 효과를 검증하였다.

Ⅲ. 연구 내용

1. 요구 분석

본 연구는 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램에 대하여 다음과 같은 내용으로 요구 분석을 하였다.

- 머신러닝에 대한 교사의 인식
- 머신러닝 교육에 대한 교육 경험
- 머신러닝 교육에 대한 교육의 필요성
- 머신러닝 교육 시 선호하는 교육 방향

위와 같은 내용으로 도내 초등학교 현직교사 100명을 대상으로 요구 분석과 관련한 설문 조사를 하였다.

2. 설문 내용

도내 초등학교 현직교사 100명을 대상으로 진행한 설문 내용을 구성할 때, 2가지 내용을 중심으로 작성하였다. 첫째, 머신러닝에 대한 기초 지식 및 인식에 대한 설문 내용을 담았다. 둘째, 머신러닝 교육에 대한 경험 여부와 머신러닝 교육을 진행할 때 선호하는 도구에 대해 설문 내용을 구성하였다. 해당 설문 내용은 [그림 Ⅲ-1]과 같다.

[교사대상 데이터 라벨링 중심의 머신러닝 교육 요구분석 설문]

- 초등학생 컴퓨팅 사고력 향상을 위한
데이터 라벨링 중심의 머신러닝 교육 설문조사

본 설문은 초등학생을 대상으로 한 데이터 라벨링 중심의 머신러닝 교육을 위한 설문입니다. 아래 설문 내용을 잘 읽어 보시고, 질문 사항에 대하여 본인의 생각과 일치하거나 유사한 내용에 번호로 표시해주세요. (예상 설문 소요시간: 2~3분 내외)

1. 소프트웨어 교육에 대하여 잘 알고 있습니까?(5에 가까울수록 잘 알고 있음)

1	2	3	4	5
---	---	---	---	---

2. 학생들을 대상으로 소프트웨어교육과 관련된 수업을 진행한 적이 있습니까?

- ① 있다. (2-1, 2-2번 문항으로)
② 없다. (2-3번 문항으로)

2-1. 소프트웨어교육에서 자주 활용하시는 방법 및 도구를 골라주세요.

- ① 언플러그드 활동 : SW보드게임, 활동지, 놀이활동 등
② 교육용 프로그래밍 언어(EPL) : 스크래치, 엔트리, 앱인벤터 등
③ 프로그래밍 언어 : 파이썬, C, 자바 등
④ 피지컬 컴퓨팅(로봇 활용) : 햄스터봇, 오조봇, 마이크로비트 등

2-2. 소프트웨어교육을 진행하실 때 가장 애로사항은 무엇입니까?

- ① 학생들 간의 성취 격차 ② 교재의 미비 ③ 관련 기자재 부족 ④ 수업 시간 부족
⑤ 기타()

2-3. 소프트웨어교육 관련 수업을 진행한다면 어떤 방법 및 도구를 사용하시겠습니까?

- ① 언플러그드 활동 : SW보드게임, 활동지, 놀이활동 등
② 교육용 프로그래밍 언어(EPL) : 스크래치, 엔트리, 앱인벤터 등
③ 프로그래밍 언어 : 파이썬, C, 자바 등
④ 피지컬 컴퓨팅(로봇 활용) : 햄스터봇, 오조봇, 마이크로비트 등

3. 머신러닝에 대하여 잘 알고 있나요?(5에 가까울수록 잘 알고 있음)

1	2	3	4	5
---	---	---	---	---

※ 머신러닝

인공지능의 한 분야로서, 사람이 학습하듯이 컴퓨터에도 데이터들을 줘서 학습하게 함으로써 새로운 지식을 얻어내게 하는 분야

[그림 III-1] 요구 분석 설문지 내용

위와 같은 내용으로 설문을 진행한 결과는 다음과 같다.

<표 III-1> 머신러닝에 대한 이해도

잘 알고 있다	대체로 알고 있다	보통이다	조금 알고 있다	전혀 모른다
23(23%)	12(12%)	33(33%)	25(25%)	7(7%)

머신러닝(인공지능)에 대해 얼마나 알고 있는지 조사한 결과는 <표 III-1>과 같다. 초등교사 중 32%가 머신러닝(인공지능)에 대해 잘 알고 있거나 대체로 알고 있다고 응답했으며, 전혀 알지 못한다고 응답한 경우는 35%로 나타났다. 이후 분석에서는 응답의 신뢰성을 위하여 머신러닝(인공지능)에 대하여 보통 이상으로 알고 있다고 응답한 65명만을 대상으로 집계하였다.

<표 III-2> 머신러닝 교육의 필요성

꼭 필요하다	필요하다	보통이다	잘 모르겠다	필요하지 않다
26(38%)	19(28%)	15(22%)	6(9%)	2(3%)

머신러닝(인공지능) 교육이 필요하다고 생각하는지 조사한 결과는 <표 III-2>와 같다. 초등교사 중 66%는 머신러닝(인공지능) 교육이 꼭 필요하다고 생각하거나 필요하다고 응답했고, 필요하지 않다고 응답한 경우는 2%에 그쳤다.

<표 III-3> 데이터 라벨링에 대한 이해

잘 알고 있다	대체로 알고 있다	보통이다	조금 알고 있다	전혀 모른다
30(44%)	16(24%)	11(16%)	10(15%)	1(1%)

데이터 라벨링에 대해 얼마나 알고 있는지 조사한 결과는 <표 III-3>과 같다. 초등교사 중 44%가 데이터 라벨링에 대해 잘 알고 있다고 응답했으며, 전혀 알

지 못한다고 응답한 경우는 1%로 나타났다.

<표 III-4> 데이터 라벨링 교육의 필요성

꼭 필요하다	필요하다	보통이다	잘 모르겠다	필요하지 않다
14(20%)	24(34%)	20(29%)	9(13%)	1(4%)

데이터 라벨링 교육이 필요하다고 생각하는지 조사한 결과는 <표 III-4>와 같다. 초등교사 중 54%는 데이터 라벨링 교육이 꼭 필요하거나 필요하다고 응답했고, 필요하지 않다고 응답한 경우는 4%에 그쳤다.

사전 요구 분석 결과 머신러닝에 대해 알고 있다고 응답한 교사들은 데이터 라벨링의 개념 역시 이해하고 있으며, 데이터 라벨링 교육의 필요성을 공감하고 있었다. 이에 본 연구에서는 데이터 라벨링 중심의 머신러닝 교육 프로그램을 개발하고, 학생들이 데이터 라벨링을 중심으로 머신러닝 프로그램을 작성하는 과정에서 컴퓨팅 사고력 향상을 꾀하는 데 중점을 두었다.

3. 교육 방향

사전 요구 분석 결과와 학습자의 수준을 고려하여 전체 12차시로 구성된 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램을 <표 III-5>와 같이 설계하였다.

<표 III-5> 교육 프로그램의 주제

차시	학습 주제
1~3	- 머신러닝 개념과 실제 서비스 소개 - 사전 검사, 데이터 레이블링 기반
4~9	- 기계 학습의 개념 배우기 - 머신러닝 포 키즈를 활용한 학습
10~12	- 조별 활동 - 사후 검사

머신러닝에 대한 개념이 초등학생들에게 어렵다는 요구 분석의 응답 결과를 고려하여, 1~3차시는 인공지능, 머신러닝, 데이터 라벨링의 개념과 실제 서비스에서 활용하는 방식을 활동지를 활용한 언플러그드 활동을 통해 학습자가 직접 찾아보고, 관련 자료를 정리하면서 능동적으로 개념을 익힐 수 있도록 구성하였다.

4~9차시는 프로젝트 학습 전략을 선정하여, 머신러닝 포 키즈를 활용한 블록형 프로그래밍 언어 기반의 교육 프로그램을 조별 활동으로 진행하면서, 데이터 라벨링 과정인 직접 데이터를 수집, 분류 및 정제하는 과정을 실습하면서, 머신러닝의 구현 원리를 실습해볼 수 있게 구성하였다. 이 과정에서 학습자가 데이터 수집, 분류, 정제 비교적 간단하게 실습할 수 있게 텍스트 타입으로 모델을 선정하고, 학습할 수 있는 문제를 선정하였다.

10~12차시는 4~9차시에서 배운 내용을 바탕으로 자율주제를 선정하여 조별 활동을 진행하도록 구성하였다. 이를 통해 적극적으로 학습에 참여하도록 지도하였다.

4. 교육 내용

프로그램 설계를 토대로 개발한 12차시의 학습 내용은 <표 III-6>과 같다.

<표 III-6> 교육 프로그램 강의 계획

차시	주제	교육 목표
1	인공지능이란 무엇인가?	1. AI의 개념을 설명할 수 있다. 2. 다양한 AI 서비스에 관해 설명할 수 있다.
2	데이터 라벨링이란 무엇인가?	1. 데이터 라벨링에 관해 설명할 수 있다. 2. 데이터의 라벨링을 어떻게 하는지 설명할 수 있다.
3	‘머신러닝 포 키즈’ 사용법	1. ‘머신러닝 포 키즈’를 사용하는 방법을 설명할 수 있다. 2. ‘머신러닝 포 키즈’에서 활용할 수 있는 머신러닝 데이터 타입과 활용 방법에 관해 설명할 수 있다.
4	신문 헤드라인 분류하기 1 - 계획하기	1. 신문 헤드라인을 분류하기 위한 머신러닝 프로그램 계획 과정에 관해 설명할 수 있다.
5	신문 헤드라인 분류하기 2 - 데이터 라벨링	1. 신문 헤드라인 수집 및 데이터 라벨링을 실행하고, 과정을 설명할 수 있다.
6	신문 헤드라인 분류하기 3 - 프로그램 만들기	1. 학습된 머신러닝을 사용하여 헤드라인을 분류하는 프로그램 코딩을 어떤 순서로 진행하는지 설명할 수 있다.
7	트위터는 어떻게 생각할까? 1 - 계획하기	1. 트위터의 타임라인 분류하는 머신러닝을 만드는 방법 설명할 수 있다.
8	트위터는 어떻게 생각할까? 2 - 데이터 라벨링	1. 트위터 타임라인 수집 및 데이터 라벨링을 수행하며 머신러닝 학습하는 과정을 실행해보고, 설명할 수 있다.
9	트위터는 어떻게 생각할까? 3 - 프로그램 만들기	1. 학습된 머신러닝 모델을 사용하여 트위터의 타임라인을 분류하는 프로그램을 코딩하고, 실행 과정을 설명할 수 있다.
10	우리가 만든 머신러닝 1 - 계획하기	1. 조별 활동으로 주제를 정하여 머신러닝 모델을 만드는 계획을 세우고, 설명할 수 있다.
11	우리가 만든 머신러닝 2 - 데이터 라벨링	1. 정해진 주제에 알맞은 데이터를 정하고, 해당 데이터를 수집한 뒤, 학습시킬 데이터를 라벨링 하는 방법을 설명할 수 있다.
12	우리가 만든 머신러닝 3 - 프로그램 만들기	1. 학습된 머신러닝 모델을 바탕으로 목표한 주제를 수행할 수 있는 프로그램을 만들 수 있다.

5. 교육자료 개발

1차시~2차시는 학습자에게 인공지능과 머신러닝의 개념을 전달하는 과정으로, 언플러그드 활동으로 구성하였고, 3차시는 4차시 이후 활용할 머신러닝 포키즈 사용법을 전달하는 과정으로 개발하였다. 1차시~2차시의 언플러그드 활동은 학습자들이 어렵다고 느낄 수 있는 인공지능 관련 개념에 대한 심리적 장벽을 낮추고, 친숙하게 느낄 수 있게 구성하는 것을 목표로 정하고, <표 III-7>과 같이 5단계 학습 과정을 구성하였다.

<표 III-7> 1차시~2차시 언플러그드 활동 계획



1단계에서는 [그림 III-2]와 같이 학습자에게 실제 생활에서 경험할 수 있는 인공지능 서비스를 소개하고, 소개한 인공지능 서비스에서 데이터 라벨링을 진행하는 방식을 소개하고, 학습된 머신러닝 모델로 서비스하는 방법을 제시하였다. 이 과정을 통해 학습자가 인공지능 기술에 대한 심리적 장벽을 낮추고, 인공지능을 이해하기 위한 배경지식을 습득할 수 있도록 구성하였다.

I. 인공지능이란 무엇인가?

학습목표 | 인공지능 정의를 이해하고, 인공지능을 활용한 서비스에 대해 설명할 수 있다.

☺ **인공지능이란?**

넷플릭스

1998년에 비디오 대여 사업부터 시작한 넷플릭스는 온라인 스트리밍을 위주로 서비스하고 있는 세계 최대의 엔터테인먼트 기업입니다. 넷플릭스는 어떻게 세계 최대의 기업이 됐을까요?



넷플릭스는 사용자가 선호하는 콘텐츠를 파악한 후 이를 바탕으로 유사한 콘텐츠를 사용자에게 추천해주는 시스템을 도입했습니다. 추천시스템이란 사용자가 일일이 검색하지 않고, 사용자의 취향에 맞는 콘텐츠를 모아 보여줌으로써 많은 사용자에게 호평 받은 서비스입니다.

[그림 III-2] 1단계 학습 과정 예시

2단계~4단계는 학습자가 학문적인 정의를 익히기 전, 학문적 정의에 사용되는 개념과 단어를 찾아보며, 용어에 익숙해지도록 구성하였다. 먼저, 2단계에서는 학습자가 인공지능 관련 개념어를 인터넷 검색을 활용하여 관련 기사나 기술 블로그, 영상 자료 등을 찾아보며, [그림 III-3]과 같이 정리하며, 인공지능 관련 개념을 이해하기 위한 배경지식을 습득할 수 있도록 구성하였다.

☺ 머신러닝이란 무엇인가?

1) '머신러닝'을 인터넷으로 검색하여, 찾은 결과를 다양한 방식으로 정리해봅시다.

1. 학습할 거리를 일단 던져 놓으면 이걸 가지고 스스로 학습하는 기계
2. 고안된 방법
3. 컴퓨터를 인간처럼 학습 시킴으로써 인간의 도움없이 컴퓨터가 스스로 규칙을 생성하는 것
4. 컴퓨터가 데이터를 통해 스스로 학습하는 것처럼 하는 기술
5. 사용자의 데이터를 기반으로 학습하거나 성능을 개선하는 시스템 인공지능

[그림 III-3] 2단계 학습 과정 실습 결과물

3단계에서는 [그림 III-4]와 같이 2단계에서 정리한 내용 중 가장 많이 나오는 단어를 빈도수를 기준으로 정리하고, 정리된 단어는 발표를 통해 학습자끼리 공유하여, 개념을 정리하는 데 많이 활용된 단어를 익히도록 구성하였다.

2) '머신러닝'을 인터넷으로 검색한 결과 중 반복되는 단어를 찾아, 몇 번 반복되었는지 정리해봅시다.

단어	횟수	단어	횟수
학습	5	스스로	3
컴퓨터	3		
인간	2		
데이터	2		

[그림 III-4] 3단계 학습 과정 실습 결과물

4단계에서는 [그림 III-5]처럼 3단계에서 빈도수로 선정된 3개 단어를 활용하여 한 줄 정의를 만들어보는 과정을 통해 학습자가 앞서 2단계~3단계를 통해 익

힌 배경지식을 상기시키고, 마지막으로 제시되는 학술적 정의를 쉽게 받아들일 수 있게 학습활동을 전개하였다.

3) 가장 많이 나온 단어 3개를 활용하여, '머신러닝'의 정의를 내려봅시다. 높, 색깔, 책

컴퓨터가 스스로 학습하여 비동적이며 알고리즘을 만든다.

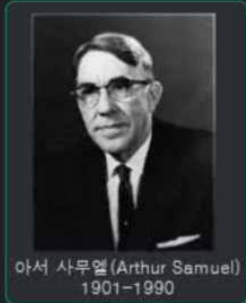
[그림 III-5] 4단계 학습 과정 실습 결과물

이후 [그림 III-6]과 같이 실제 학문적 개념을 제시하여, 학습자가 조사하여 정리한 결과와 비교하여 학술적인 개념을 익힐 수 있게 진행하였다.

머신러닝이란 무엇인가?

**컴퓨터가
명시적으로
프로그램되지 않고도
학습할 수 있도록 하는 연구 분야**

(아서 사무엘, Arthur Lee Samuel, 1959)

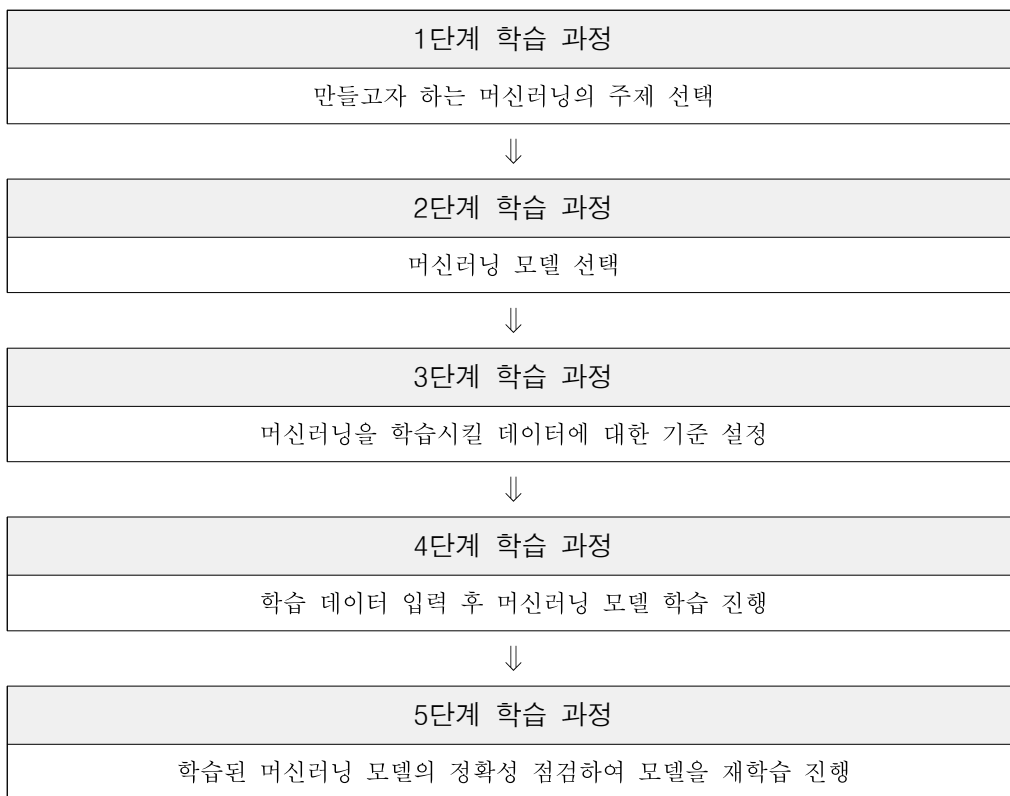


아서 사무엘 (Arthur Samuel)
1901-1990

[그림 III-6] 5단계 학습 과정 예시

4차~9차시는 머신러닝 포 키즈의 프로젝트를 활용하여 <표 III-8>에 제시한 5 단계를 순차적으로 진행하도록 교육자료를 개발하였다. 4차시~9차시에서 학습자가 단계별로 머신러닝 주제 선정, 주제에 따른 머신러닝 모델 선정, 모델에 따른 데이터 라벨링 기준 선정 등을 조별 활동으로 진행하며, 머신러닝 모델을 만들어 가는 과정을 단계적으로 체험하고 익힐 수 있도록 구성하였다.

<표 III-8> 텍스트 데이터 라벨링 실습 활동 계획



이 활동에서는 머신러닝 포 키즈에 제시된 프로젝트를 활용하되 [그림 III-7]에 제시한 활동지를 통해, <표 III-8>로 제시한 텍스트 데이터의 데이터 라벨링을 실습 5단계를 실행할 수 있게 구성하였다.

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝포키즈에서 '워크시트' 메뉴에서 선택합니다.)

[학습 데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터종류	
데이터 라벨	
라벨 별로 데이터 입력 기준 만들기 (예시 포함)	
최소로 입력할 데이터 수	

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝포키즈' 프로젝트에 데이터를 입력합니다.

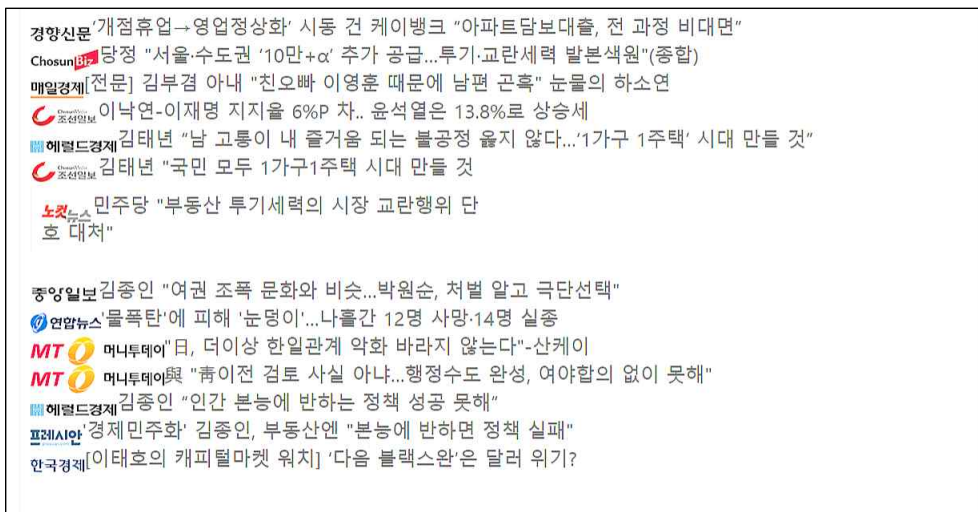
[훈련된 모델 정확도 테스트 하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도

[그림 III-7] 텍스트 데이터 라벨링 활동지

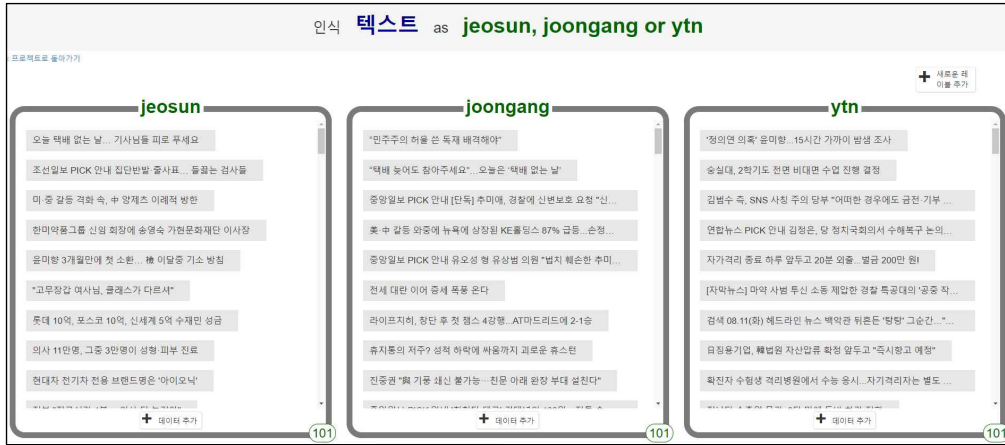
학습자는 [그림 III-7]에 제시한 활동지의 구성에 따라 해결하고 하고자 하는 문제와 관련된 텍스트 데이터를 수집하는 방법에 대해 학습자가 조별 활동을 통해 기준을 정하고, 정해진 기준에 따라 데이터를 직접 수집하여, 수집한 데이터를 살펴보고, 학습할 수 있는 데이터의 형태로 변경하여, 데이터 라벨링을 실습하는 과정을 구성하였다.

이는 선행 연구된 인공지능 교육 프로그램에서 기존의 검증된 데이터를 활용하는 것과 차별화된 단계로서 학습자가 직접 학습데이터를 수집하고, 데이터 라벨링을 진행하는 실습을 추가하여 구성하였다. 또한 훈련한 모델을 평가하는 활동을 넣어, 모델의 정확도를 높이기 위해 데이터 라벨링 과정의 성찰하고, 재학습시키는 교육활동을 구성하였다.



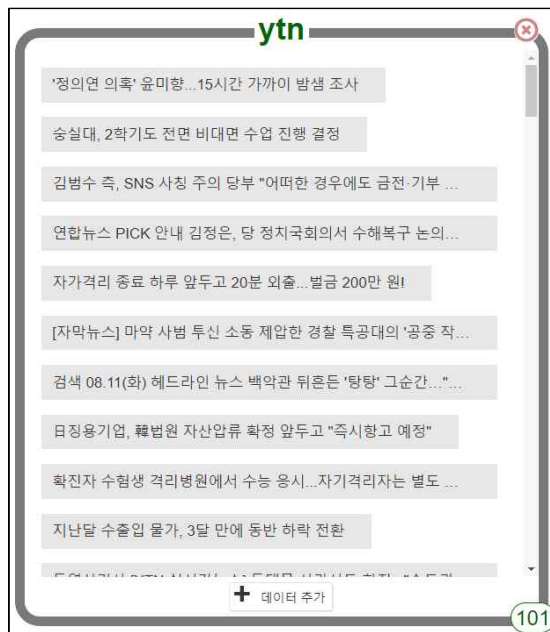
[그림 III-8] 텍스트 데이터 수집 예시

[그림 III-7]에 제시된 활동지에 따라 학습자들은 [그림 III-8]과 같이 인터넷 검색을 통해 학습과 관련한 텍스트 데이터를 수집하였다. 초기 데이터 수집 단계에서는 모델을 학습시킬 때 꼭 필요한 텍스트 정보만 제시하고, 최대한 있는 그대로의 데이터를 수집할 수 있도록 지도하였으며, 해당 데이터에 대한 분별없이 주어진 시간 내에 최대한 많은 텍스트 데이터를 수집할 것을 전달하였다.



[그림 III-9] 텍스트 데이터 라벨링 예시

학습자들은 [그림 III-8]과 같이 수집한 텍스트 데이터를 분석하여, 활동지를 활용하여, [그림 III-9]와 같이 텍스트 데이터를 분류할 라벨링 기준을 선정하였다. [그림 III-9]의 경우, [그림 III-8]에 수집한 텍스트 데이터 중 해당 데이터를 작성한 언론사를 기준으로 분류한 실습 결과를 제시한 결과물이다.



[그림 III-10] 텍스트 데이터 라벨링에 따른 데이터 입력 결과 예시

학습자는 활동지를 활용하여 텍스트 데이터를 분류하는 기준을 세우고, 수집한 텍스트 데이터를 정제하여, [그림 III-10]처럼 학습 데이터로 입력하는 과정을 실습할 수 있도록 진행하였다. 이 과정에서 학습자들은 인터넷 검색을 통해 찾은 텍스트 데이터를 입력하기 편하게 태그를 제거하는 방법을 익히고, 입력하면서 중복 데이터가 입력될 때, 나타나는 에러 메시지를 확인하는 방법 등을 실습하며 데이터 전처리 과정을 간략하게 체험해 볼 수 있도록 구성하였다.

10~12차시는 1~9차시의 학습 내용을 바탕으로 학습자들이 조별 활동을 통해 직접 머신러닝 주제를 선택하고, 주제에 맞는 모델을 선택하여 데이터 라벨링을 어떻게 진행할 것인지 계획을 세우고, 계획에 따라 데이터를 수집하고, 정제하여, 분류하는 과정을 거쳐 학습 모델을 만들어 블록형 프로그래밍 언어로 프로그램을 만들어보는 차시로 구성하였다.

IV. 연구 결과 및 해석

1. 연구 대상

본 연구는 제주시 OO초등학교에 재학 중인 6학년 학생 17명을 대상으로 진행하였다. 참여한 학생들은 스크래치와 엔트리를 활용한 소프트웨어 교육을 받은 경험이 있지만, 인공지능에 대한 교육 경험이 없었다.

2. 검사 도구

본 연구에서는 컴퓨팅 사고력을 측정하는 도구로 비버 챌린지 한국 그룹 III의 도전하기 문항(비버 챌린지, 2018)과 김병수의 연구에서 개발한 계산적 인지력 검사(Computational Cognition Test) A, B형을 사용하였다(김병수, 2014). 이 두 가지 검사를 활용하여, 텍스트 데이터의 데이터 라벨링을 활용한 머신러닝 교육 프로그램의 효과성에 대한 측정을 진행하였다.

3. 연구 설계 및 처치

데이터 라벨링 중심의 머신러닝 교육을 하기 전에 비버 챌린지와 계산적 인지력 검사 A형을 활용하여 사전 검사를 하였다. 이후 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램을 6주간 총 12차시의 내용으로 구성하여, 진행하였다.

교육 종료 후 실험집단의 교육 효과를 검증하기 위해 비버 챌린지와 계산적 인지력 검사 B형을 활용하여 사후 검사를 하였으며, 본 연구의 설계를 도식화하여, <표 IV-1>에 제시하였다. 통계 분석은 IBM SPSS Statistics Subscription (11-2018)를 활용하여 진행하였다.

<표 IV-1> 실험 설계

대상	사전 검사	적용	사후 검사
G	O ₁	X	O ₂
	O ₃		O ₄

G : 실험집단 (N = 17)

O₁, O₂ : 비버 챌린지

O₃ : 계산적 인지력 검사 A형

O₄ : 계산적 인지력 검사 B형

X : 텍스트 테이더 라벨링 기반 머신러닝 교육 프로그램

4. 사전-사후 검사의 정규성 검정

실험집단의 표본의 크기가 17명으로 다소 적어, 실험집단의 컴퓨팅 사고력의 사전 검사와 사후 검사 결과가 정규성을 만족하는지 확인하기 위해 정규성 검정을 하였다. 정규성 검정은 보편적으로 활용하는 Shapiro-Wilks 검사를 활용하였으며, 해당 결과는 <표 IV-2>, <표 IV-3>과 같다.

<표 IV-2> 비버 챌린지 정규성 검정

검사 기간	설명 통계(N=17)				통계값	유의확률
	평균	표준편차	최대값	최소값		
사전 검사	18.000	22.226	66.00	-18.00	.955	.538
사후 검사	32.352	22.363	82.00	-6.00	.963	.690

*p<.05

<표 IV-3> 계산적 인지력 검사 정규성 검정

검사 기간	설명 통계(N=17)				통계값	유의확률
	평균	표준편차	최대값	최소값		
사전 검사	3.1765	2.48081	9.00	.00	.911	.103
사후 검사	6.3529	3.46304	11.00	1.00	.894	.054

*p<.05

정규성 검정 결과 비버 챌린지와 계산적 인지력 검사 A, B형의 사전 검사 결과와 사후 검사 결과가 모두 정규성을 만족한 것으로 나타났다.

5. 컴퓨팅 사고력 사전-사후 검사 비교

사전-사후 검사 결과 Shapiro-Wilks 검사에서 정규성을 확보하여, 대응 표본 t 검정을 시행하였다. 대응 표본 t 검정 결과는 <표 IV-4>에 제시하였다. 먼저 비버 챌린지의 대응 표본 t 검정 결과를 살펴보면 다음과 같다.

<표 IV-4> 비버 챌린지 대응 표본 t 검정 결과

인원	사전 검사		사후 검사		통계값	유의확률
	평균	표준편차	평균	표준편차		
17	18.000	22.226	32.352	22.363	-3.727	.002**

*p<.05 **p<.01

‘컴퓨팅 사고력’의 평균 점수는 사전 18.000에서 사후 32.352로 14.352점 상승하였고, 유의확률은 .002로 평균 점수가 통계적으로 유의미하게 향상된 것으로 나타났다.

비버 챌린지와 함께 측정한 계산적 인지력 검사의 대응 표본 t 검정 결과는 <표 IV-5>와 같다.

<표 IV-5> 계산적 인지력 검사 대응 표본 t 검정 결과

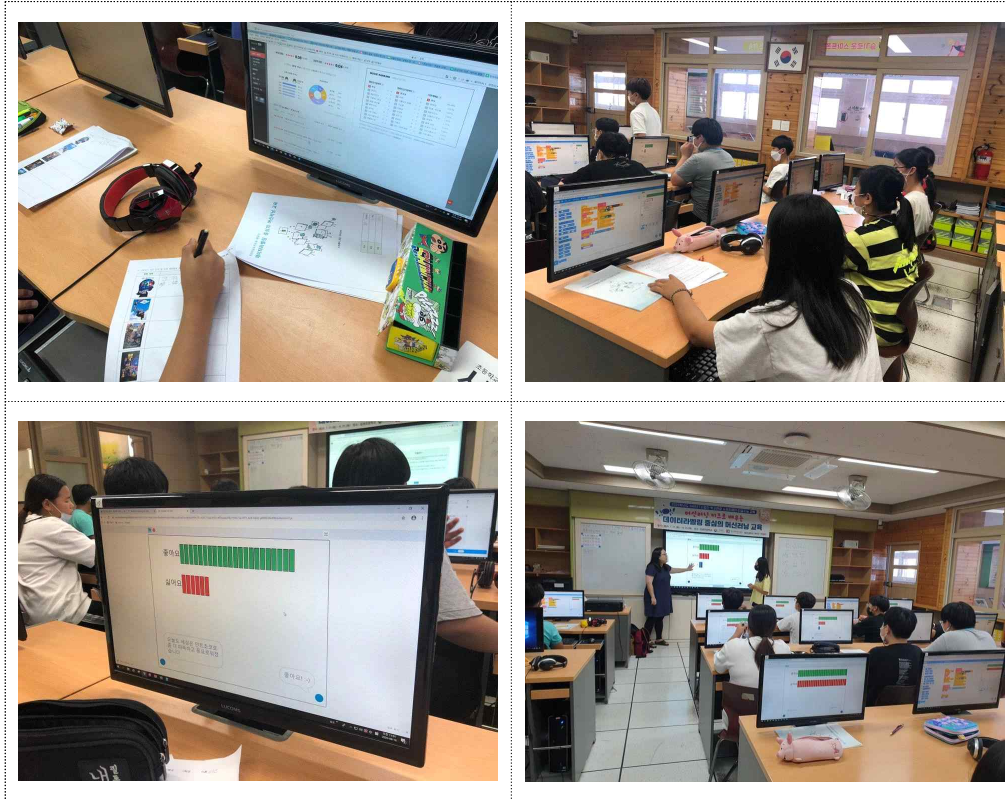
인원	사전 검사		사후 검사		통계값	유의확률
	평균	표준편차	평균	표준편차		
17	3.1765	2.48081	6.3529	3.46304	-5.700	.000**

*p<.05 **p<.01

계산적 인지력 검사의 평균 점수는 사전 3.1765에서 사후 6.3529로 3.1764 점 상승하였고, 유의확률은 .000로 평균 점수가 통계적으로 유의미하게 향상된 것으로 나타났다.

6. 학생 산출물 심층 분석

본 연구에서 개발한 데이터 라벨링 중심의 머신러닝 교육 프로그램은 OO초등학교 6학년을 대상으로 6주에 걸쳐 총 12차시에 걸쳐 진행되었다. 학생들은 [그림 IV-1]과 같이 1~3차시까지 언플러그드 활동을 진행하고(왼쪽 위), 4~9차시까지 활동지를 활용하여, 텍스트 데이터 수집과 라벨링을 진행하는 활동을 했다. 이후 10~12차시에는 1~9차시까지 배운 내용을 바탕으로 조별 프로젝트를 진행하며, 조에서 결정한 주제에 관련한 텍스트 데이터를 수집하고, 해당 데이터를 바탕으로 모델을 학습시키고, 정확도를 확인하는 작업을 진행하였다.



[그림 IV-1] 머신러닝 교육 프로그램 수업 장면

학생들의 4~9차시의 산출물은 [그림 IV-2]와 [그림 IV-3]과 같다. 10~12차시의 산출은 [그림 IV-4]와 [그림 IV-5]와 같다.

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

허리안을 리어, 산문을 맞추는 인공지능

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝포키즈에서 '워크시트' 메뉴에서 선택합니다.)

히도인 테스트

[학습 데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터종류	텍스트
데이터 라벨	진, 중, 안
라벨 별로 데이터 입력 기준 만들기 (예시 포함)	
<p>신문가 모은 방법</p> <p>1) 네이버에서 기사를 검색한다</p> <p>2) 검색한 기사에서 히도인을 찾아서 붙여 붙인다</p> <p>3) 신문사가 맞췄다 여부를 입력한다</p> <p>주의사항: 다른 신문사와 비교하여 비슷한 허리안을 10번 이상 넣는다</p>	
최소 입력할 데이터 수	20

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝포키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트 하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도
45%	87%			

[그림 IV-2] 4~9차시 활동지 예시 1

◎ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

뉴스 인문사... / 어떤 인문사의 신간이인지 찾기

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.

(머신러닝포키즈에서 '워크시트' 메뉴에서 선택합니다.)

헤드라인 테스트

[학습 데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터종류	텍스트 [Text]
데이터 라벨	중앙/조선 YTN
라벨 별로 데이터 입력 기준 만들기 (예시 포함)	
1. 뉴스키 들어가다 (비버) 2. 인문사뉴스 / 3. 헤드라인 따고 복붙 4. 신문사끼리 맞춰 데이터 입력	
최소로 입력할 데이터 수	20개

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝포키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트 하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도
100%	100%	100%	100%	100!!!

95%

100%
100%
완전합치

[그림 IV-3] 4~9차시 활동지 예시 2

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

(귀여)를 검색할때 나는 다양한 공식인지 발견하지 못함

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝포키즈에서 '워크시트' 메뉴에서 선택합니다.)

트위터는 어떻게 생각하십니까?

[학습 데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터종류	트윗
데이터 라벨	좋아, 싫어, 몰라
라벨 별로 데이터 입력 기준 만들기 (예시 포함)	
<p>트윗의 분류는 3가지</p> <ol style="list-style-type: none"> 1) '좋아, 기쁘게함' 등의 단어가 들어가면 좋아 이분다 2) 나쁜말(독어 부정적인 말)이 들어가면 싫어 이분다 3) 같은 단어들끼리만 몰라 이분다 <p>주제명: 다양한 검색어 행이나 양도 heilker 이분다</p>	
최소 입력할 데이터 수	30개

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝포키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트 하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도
80%	85%	90%	95%	98%

[그림 IV-4] 10~12차시 활동지 예시 1

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

팬츠(X)를 선택했을 때 보는 타임라인이 긍정/부정인지 분류.

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝포키즈에서 '워크시트' 메뉴에서 선택합니다.) 링트족호 ♡

트위터는 어떻게 생각하나요??

[학습 데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터종류	텍스트
데이터 라벨	like like (좋음) / dislike (싫음)
라벨 별로 데이터 입력 기준 만들기 (예시 포함)	
<p>1 like : 사랑, 좋아, <u>호</u>, <u>관</u>, <u>최고</u>, 좋아 았어</p> <p>2 dislike : 싫어, <u>불호</u>, <u>분</u>, <u>나쁜말</u> XX ii, <u>안 좋아</u></p> <p>3 <u>좋은</u> <u>감</u> <u>증거</u> 'like' / <u>나쁜말</u> dislike 감 <u>부름</u></p>	
최소로 입력할 데이터 수	1개 4개

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝포키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트 하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도
100%	80%	65%	91%	90%
99%	100%	100%	62%	88%

[그림 IV-5] 10~12차시 활동지 예시 2

각 산출물을 비교해 보면 4차시~9차시보다 10차시~12차시에서 선정 기준이 더 구체적이고, 데이터 수집 시 기준이 되는 단어를 선정하는 등 데이터 수집-정제-분류를 고려한 기준이 추가되었음을 확인할 수 있다. 또한 학습자가 학습에 필요한 단어 수를 최소 10개에서 최대 20개까지 늘어나 학습하는 데이터의 개수가 많을수록 정확한 결과가 나온다는 것을 체득하였음을 알 수 있었다.

그리고 학습 경험이 쌓일수록 모델을 생성한 후 생성된 모델의 정확도를 측정하는 횟수와 측정 이후 정확도를 향상한 기록이 늘어나고 있음을 알 수 있다. 이를 통해, 학습자가 데이터 라벨링을 통해 모델의 정확도를 높일 수 있다는 것을 인식하고, 모델의 성능을 높이기 위해 수정작업을 진행한다는 것을 알 수 있었다.

7. 연구 결과 분석

텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램을 적용하기 전 비버 챌린지와 계산적 인지력 검사 A형을 활용하여 사전 검사를 진행하였다. 그리고 해당 교육 프로그램을 적용한 뒤 비버 챌린지와 계산적 인지력 검사 B형으로 사후 검사를 진행하여 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램 적용 후 실험 집단의 컴퓨팅 사고력 향상 여부를 측정하였다.

사전·사후 검사 분석에 앞서 실험집단이 정규분포를 갖췄는지 확인하기 위해 Shapiro-Wilks 검사를 시행하였다. 정규성 검정 결과 비버 챌린지의 사전 검사의 유의확률이 .538, 사후 검사의 유의 확률이 .690로써 .05보다 커 통계적으로 정규성을 만족하였다. 또한 함께 실시한 계산적 인지력 검사 A형과 B형의 정규성 검정 결과도 사전 검사의 유의확률이 .103, 사후 검사의 유의확률이 .054로 .05보다 커 정규성을 만족하는 것으로 나타났다.

Shapiro-Wilks 검사를 바탕으로 정규성을 확보하여, 비버 챌린지와 계산적 인지력 검사 A, B형으로 진행한 사전·사후 검사의 대응 표본 t 검정을 진행하였다. 대응 표본 t 검정 결과 비버 챌린지와 계산적 인지력 검사로 측정할 수 있는 학습자의 컴퓨팅 사고력이 사후 검사 결과는 사전 검사 대비하여 평균 점수가 향상되었으며, 유의확률도 통계적으로 유의미한 향상을 보였다.

또한 학생들의 산출물을 심층 분석한 결과 텍스트 데이터를 수집할 때 기준이 되는 키워드를 선정하거나, 수집하는 텍스트의 기준을 전략적으로 선정하는 등의 변화를 확인할 수 있었다. 이와 같은 학습활동의 결과물을 살펴볼 때, 데이터 라벨링에 중심을 둔 교육을 통해 컴퓨팅 사고력 향상에 긍정적인 영향을 미치는 것으로 보인다.

따라서 본 연구는 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램이 스크래치나 엔트리 등을 활용하여 소프트웨어 교육을 받은 경험이 있는 초등학생의 컴퓨팅 사고력 향상에 긍정적인 영향을 미칠 수 있다는 결과를 보여주었다. 더불어 학생들이 접근하기 쉬운 텍스트 데이터를 활용하여 자료를 수집하고, 분석하여 실제 데이터 라벨링을 진행하는 실습이 머신러닝 교육에서 가능성을 제시하였다.

V. 결론 및 제언

1. 결론

머신러닝은 4차 산업 혁명 시대의 주요한 기술의 하나로서, 사회 전반의 변화를 끌어내며, 많은 분야에 적용되고 있다. 우리 생활에서 쉽게 접할 수 있는 추천 시스템을 비롯하여 질병과 관련한 신약의 개발이나 맞춤형 헬스케어 등의 의료 서비스, 자율주행 자동차 등의 기계를 제어하는 시스템을 넘어, 그림이나 글을 생성해 내는 서비스까지 빠르게 발전하면서 광범위한 분야에 영향을 미치고 있다. 이러한 사회의 흐름에 따라 머신러닝과 관련한 교육이 실제 교육 현장에서도 관련 교육 프로그램 및 교재 개발이 이루어지고 있다.

본 연구는 이러한 사회적 흐름을 반영한 선행연구를 바탕으로 현재 교육 현장에서 운영되고 있는 2015년 개정 교육 과정 내에서 머신러닝 교육을 진행할 수 있는 머신러닝 교육 프로그램을 개발하고 적용하여 초등학생들의 컴퓨팅 사고력에 미치는 영향을 파악하고자 하였다. 이를 위해 ADDIE 모형의 절차에 따라 사전 요구 분석 결과를 바탕으로 텍스트 데이터를 중심으로 진행하는 머신러닝 교육 프로그램을 개발하였고, 6주에 걸쳐 총 12시간 동안 텍스트 데이터의 데이터 라벨링에 중점을 둔 머신러닝 교육 프로그램을 운영하였다.

교육 프로그램을 운영할 때, 학생들이 자료를 수집하고, 분석하는 과정에서 활동지를 통해, 자료의 특성과 수집 기준 및 데이터 라벨링에 대한 기준을 세울 수 있게 길잡이를 제시하여, 데이터 라벨링에 대한 개념과 활동을 학습할 수 있도록 구성하였다. 또한 학습자가 학습한 내용을 바탕으로 직접 데이터의 범위를 선정하고, 수집하는 과정을 발표할 수 있도록 구성하여, 컴퓨팅 사고력에 영향을 미칠 수 있도록 구성하였다. 교육 전후에 사전·사후 검사를 통해 컴퓨팅 사고력의 변화를 분석하였다.

본 연구를 통해 다음과 같은 결론을 도출할 수 있었다.

첫째, 스크래치나 엔트리를 경험한 학습자를 대상으로 머신러닝 교육을 진행할 경우, 언플러그드 활동으로 실제 생활에서 머신러닝이 활용되는 서비스를 찾아보

고, 직접 개념을 정리하고 확인하는 과정을 진행할 때, 이후 머신러닝 포 키즈를 활용하여 머신러닝의 모델을 학습시키고, 정확도를 평가하는 활동에 대해 중요성을 인식하고 활동하는 것으로 생각한다.

둘째, 머신러닝 교육을 접하지 않은 학습자에게 직관적이고, 판단하기에 비교적 쉬운 텍스트 데이터를 활용하여 머신러닝 모델의 성능에 직접적으로 영향을 미치는 데이터 라벨링에 대한 개념을 전달하고, 실제 학습을 통해 스스로 데이터 라벨링에 대한 계획을 수립하여, 자료를 수집하고, 분석하여 라벨을 정하는 과정을 수행할 수 있음을 알 수 있었다. 또한 데이터 라벨링 과정이 학습된 모델에 미치는 영향을 미침을 인지하고, 데이터를 추가로 수집하고, 다시 모델을 학습시키는 과정을 학습하기에 텍스트 데이터가 적합하다고 판단된다.

셋째, 본 연구에서 개발한 텍스트 데이터의 데이터 라벨링을 활용한 머신러닝 교육 프로그램의 사전 검사 결과와 사후 검사 결과를 분석하였을 때, 학습자의 컴퓨팅 사고력이 향상되는 결과를 보였다. 이는 텍스트 데이터를 활용한 데이터 라벨링 과정에서 자료 수집 및 분석을 실습해 볼 수 있고, 이후 모델 학습 과정과 학습된 모델을 스크래치를 활용하여 블록 코딩을 하는 과정에서 자동화와 추상화를 경험해 볼 수 있기 때문이라고 생각한다. 특히 자료 수집 및 분석 과정은 스크래치를 활용한 블록 코딩에서는 쉽게 접하기 어려운 경험이기 때문에, 텍스트 데이터를 활용한 데이터 라벨링을 진행하는 과정에서 습득하기 적절한 구성요소로 생각한다.

2. 제언

본 연구는 기존에 구성된 데이터를 제공하여 진행한 머신러닝 교육과 달리 학습자가 직접 자료를 수집하고, 해당 데이터를 가공하는 데이터 라벨링 과정을 실습해볼 수 있다는 것에 의의가 있다. 다만, 본 연구의 실험집단은 일반적인 상관관계 연구에 일반화시키기 다소 적은 17명으로 진행되었기 때문에 연구 결과를 일반화하는 데 한계가 있다. 더불어 본 연구에서 비교집단이 없이 실험집단의 사전·사후 검사를 활용하여 검증하였기 때문에, 본 프로그램의 영향 때문인지 상관관계를 분명하게 규정할 수 없는 한계점이 있다.

또한 다양한 데이터 형태가 활용되는 머신러닝 분야에서 텍스트 데이터로 제한하여 접근하여, 이미지 데이터, 소리 데이터 등을 활용하기에는 다소 어려움이 예상된다.

따라서 이러한 한계점을 바탕으로 추후 연구에서는 다수의 참가자를 확보하고, 실험집단과 비교집단을 구성하여 연구 결과에 대한 각 요인을 좀 더 체계적으로 분석할 필요가 있다. 또한 이번 연구에서 다루지 않았던 이미지 데이터와 소리 데이터 등 다른 데이터 형식에 따른 데이터 라벨링을 활용한 머신러닝 교육 프로그램을 개발하여, 각 데이터에 적절한 자료 수집 방법과 데이터 라벨링의 기준을 수립하는 과정을 학습자가 경험할 수 있도록 연구 분야의 확장이 필요하다.

참 고 문 헌

- 과학기술정보통신부. (2019). 과학기술정보통신부 업무보고.
- 과학기술정보통신부. (2020). 인공지능(AI) 데이터 품질 표준안, 국내외 표준화 추진 보도자료.
- 관계부처 합동. (2020). 전 국민 AI·SW 교육 확산 방안.
- 교육부. (2015). 교육부 소프트웨어 교육 정책.
- 한국 정보과학교육연합회. (2019). 차세대 인공지능 융합형 정보 인재 양성을 위한 소프트웨어 교육 혁신 포럼.
- 김갑수, 구덕희, 김성백, 김수환, 김영식, 김자미, 김재현, 김창석, 김철, 김한일, 김현철, 박남제, 박정호, 박관우, 서인순, 서정연, 성영훈, 송태욱, 이영준, 이재호, 이정서, 이현아, 이형욱, 전수진, 전용주, 정영식, 정인기, 최숙영, 최정원, 한선관. (2020). 차세대 소프트웨어(SW) 교육 표준모델 개발. **정보교육학회논문지**, 24(4), 337-367.
- 김갑수, 박영기. (2017). 초등학생의 인공지능 교육을 위한 교수학습 모델 개발 및 적용. **정보교육학회논문지**, 21(1), 139-149.
- 김갑수. (2019). 초등교사들을 위한 인공지능 교육 프로그램 개발 및 적용. **정보교육학회논문지**, 23(6), 629-637.
- 김병수. (2014). 계산적 사고력 신장을 위한 PPS 기반 프로그래밍 교육 프로그램. 박사학위논문, 제주대학교, 제주.
- 김은지, 이태욱.(2018). 컴퓨팅 사고력 평가 도구로써 비버 챌린지 2017 신뢰도 분석 : 한국 그룹 III를 중심으로. **2018년 한국컴퓨터교육학회 하계 학술발표논문지**, 22(2), 103-106.
- 류미영, 한선관(2019). 딥러닝 개념을 위한 인공지능 교육 프로그램. **정보교육학회논문지**, 23(6), 583-590.
- 문우중, 김봄솔, 김정아, 김봉철, 서영호, 오정철, 김용민, 김종훈. (2021). 데이터 라벨링 중심의 머신러닝 교육이 초등학생 컴퓨팅 사고력에 미치는 효과. **정보교육학회논문지**, 25(2), 327-335.
- 박선주.(2019). 알고리즘 교육을 통한 비버 챌린지 결과 분석. **정보교육학회논문**

- 지, 23(1), 65-72.
- 오니시 가나코. (2019). **가장 쉬운 인공지능 AI 입문서**. 서울: 아티오.
- 이보배. (2020). **데이터 불균형이 머신러닝 모델 성능에 미치는 영향**. 석사학위 논문, 고려대학교 컴퓨터정보통신대학원, 서울.
- 이영준, 백성혜, 신재홍, 유현창, 정인기, 안상진, ... 정성균. (2014). **초중등단계 Computational Thinking 도입을 위한 기초 연구**. 서울: 한국과학창의재단.
- 이영호. (2020). **모두의 인공지능 with 스크래치 - 누구나 쉽게 만들면서 배우는 인공지능 기초**. 서울: 길벗.
- 이영호. (2019). 블록형 프로그래밍 언어 기반 인공지능 교육이 학습자의 인공지능 기술 태도에 미치는 영향 분석. **정보교육학회 논문지**, 23(2), 189-196.
- 이은경. (2020). 국내외 초·중등학교 인공지능 교육 과정 분석. **한국컴퓨터교육학회 논문지**, 23(1), 37-44.
- 위키백과. (2013). 인공지능. (n.d.). 2021. 1. 13. <https://ko.wikipedia.org/wiki/>
- 장준혁. (2018. 8. 7). 새로운 인공지능 기술 GAN ① - 스스로 학습하는 인공지능. 2021. 1. 13. <https://www.samsungsds.com/kr/insights/Generative-adversarial-network-AI.html>
- 정영식, 김갑수, 정인기, 김현배, 김철, 유정수, 김종우, 홍명희. (2015). 초등학생을 위한 SW 교육과정 표준 모델 개발. **한국정보교육학회 논문지**, 19(4), 467-480.
- 한국정보화진흥원. (2019). **글로벌 인공지능 연구의 4대 키워드와 시사점-4대 키워드 : 인간, 유용, 안전, 이해-**. 서울:한국정보화진흥원.
- Bebras Challenge. (n.d.). *Introduction to Bebras Challenge*. Retrieved March 10, 2021, from <https://bebraschallenge.org/>
- Lou C, Walter D. (2009). **체제적 교수 설계**, 서울: 아카데미프레스.
- Machine Learning For Kids(n.d.). *Introduction to Machine Learning For Kids*. Retrieved March 10, 2021, from <https://machinelearningforkids.co.uk/>
- Margaret R. (n.d.). *Definition of Data labelin*. Retrieved March 10, 2021, from

<https://whatis.techtarget.com/definition/data-labeling>

ISTE. (n.d.). *Introduction to ISTE*. Retrieved March 10, 2021, from <https://www.iste.org/>

Wing, J. M. (2006). Computational Thinking. *Communication of the ACM*, 49(3), 33-35.

A B S T R A C T *

Development and application of the Machine Learning educational program based on text data labeling

Kim, Bom Sol

Major in Elementary Computer Education
Graduate School of Education
Jeju National University

Supervised by Professor Kim, Jong Hoon

This study developed and applied the text data labeling-based Machine Learning education program, as an educational method to improve the computational thinking of elementary school students, and then, verified its effectiveness. To develop the educational program, for 100 elementary school teachers in Jeju Island, their needs and requirements were analyzed in advance and based on such results, the program was designed and developed. In the developed educational program, 12 sessions were held for 6 weeks, targeting 17

* A thesis submitted to the committee of Graduate School of Education, Jeju National University in partial fulfillment of the requirements for the degree of Master of Education conferred in May, 2021.

male and female students enrolled in the 6th grade of the elementary school. The effectiveness of the program was verified in the following ways; the pre- and post-tests were conducted using 'Bebras Challenge' and 'Computational Cognition Test Type A and B' in order to analyze the educational effects. Such analyses showed that Machine Learning education focused on data labeling contributed to the improvement of elementary school students' computational thinking.

Key words: text data, data labeling, Machine Learning Education, Machine Learning for kids, Computational thinking

부 록

[부록 1] 텍스트 데이터 라벨링 기반 머신러닝 교육 강의 계획안

[부록 2] 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램 교재

[부록 1]

인공지능(머신러닝) 강의 계획안(12차시)

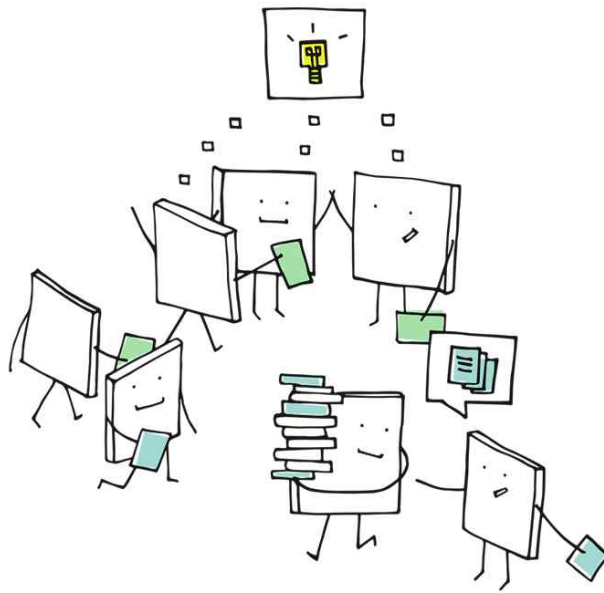
프로그램명	인공지능(머신러닝)	지도 강사	김봄솔
대 상	초등학교 6학년	강의실	컴퓨터실
지도 기간	7월 21일 ~ 8월 25일	지도 시간	화요일 10 : 30~12 : 00
지도목표	4차 산업혁명 시대를 이끄는 대표기술인 인공지능(머신러닝)의 개념을 익히고, 인공지능(머신러닝)을 데이터 모으기, 데이터 학습시키기, 학습시킨 결과를 바탕으로 스크래치 3을 활용한 프로젝트를 만들어 컴퓨팅 사고력을 향상한다.		
차시	학습 주제	학 습 목 표	
1차시	인공지능이란?	<ol style="list-style-type: none"> 1. 인공지능의 개념과 역사를 살펴보고, 인공지능에 대해 설명할 수 있다. 2. 다양한 인공지능 서비스를 체험해보고, 인공지능으로 어떤 서비스를 만들 수 있는지 설명할 수 있다. 	
2차시	데이터 라벨링이란?	<ol style="list-style-type: none"> 1. 인공지능 학습을 위해 필요한 데이터 라벨링이 어떤 일인지 설명할 수 있다. 2. 인공지능 학습을 위한 데이터 라벨링을 하는 방법을 수행할 수 있다. 	
3차시	'머신러닝 포 키즈' 활용하기	<ol style="list-style-type: none"> 1. 인공지능의 한 분야인 머신러닝을 실제로 실습할 때 필요한 '머신러닝 포 키즈'를 활용할 수 있다. 2. '머신러닝 포 키즈'에서 제시하는 머신러닝의 종류와 다양한 프로젝트 활용 방식을 설명할 수 있다. 	
4차시	신문 헤드라인 정복하기1 -계획하기	<ol style="list-style-type: none"> 1. 신문의 헤드라인을 보고, 어떤 신문의 헤드라인인지 맞출 수 있는 머신러닝을 만드는 방법을 설명할 수 있다. 	

5차시	신문 헤드라인 정복하기 2 -데이터 라벨링 하기	1. 실제 신문사 헤드라인을 수집하고, 인공지능 학습을 진행할 수 있게 정리하는 데이터 라벨링을 z 수행할 수 있다.
6차시	신문 헤드라인 정복하기 3 -프로그램 만들기	1. 학습이 완료된 머신러닝을 활용하여, 실제 기사의 헤드라인을 신문사별로 분류하는 프로그램을 만들 수 있다.
7차시	트위터는 어떻게 생각하나요1 -계획하기	1. 트위터에 검색한 주제를 스스로 정한 기준에 맞춰 분류하는 머신러닝을 만드는 방법을 설명할 수 있다.
8차시	트위터는 어떻게 생각하나요2 -데이터 라벨링 하기	1. 주제를 선정하고, 실제 트위터의 타임라인을 수집하여, 인공지능 학습을 진행할 수 있게 정리하는 데이터 라벨링을 수행할 수 있다.
9차시	트위터는 어떻게 생각하나요3 -프로그램 만들기	1. 학습이 완료된 머신러닝을 활용하여, 실제 트위터의 타임라인을 분류하는 프로그램을 만들 수 있다.
10차시	우리가 만든 인공지능1 -계획하기	1. 모둠 활동을 하면서, 텍스트 데이터를 학습한 결과를 바탕으로 실행되는 머신러닝 프로그램을 만드는 계획을 세울 수 있다.
11차시	우리가 만든 인공지능2 -데이터 라벨링 하기	1. 모둠 활동으로 주제에 맞게 자료를 수집하고, 분류하는 작업을 수행할 수 있다.
12차시	우리가 만든 인공지능3 -프로그램 만들기	1. 학습이 완료된 머신러닝을 활용하여, 모둠별 계획에 맞춰 프로그램을 만들 수 있다.

[부록 2]

머신러닝포키즈로 배우는

데이터라벨링 중심의 머신러닝 교육



학교	초등학교
학년	학년
이름	

I. 인공지능이란 무엇인가?

학습 목표 | 인공지능 정의를 이해하고, 인공지능을 활용한 서비스에 대해 설명할 수 있다.

☺ 인공지능이란?

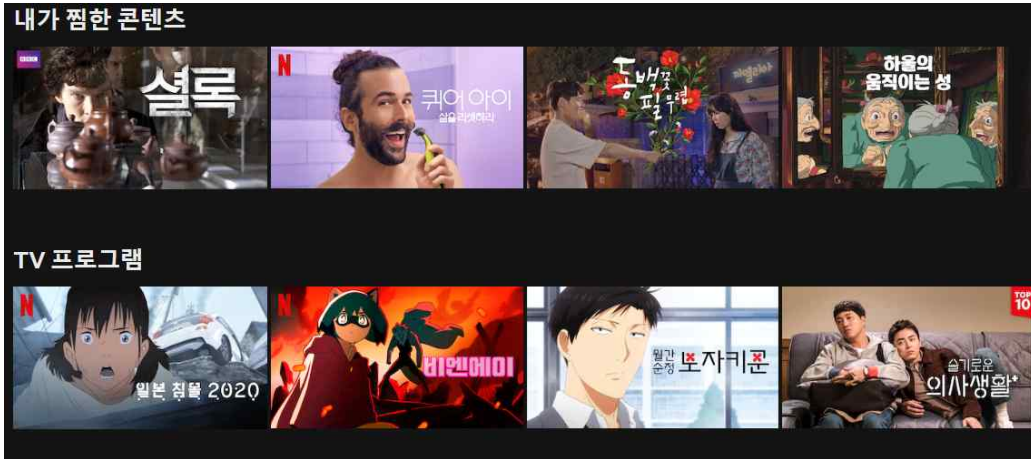
넷플릭스

1998년에 비디오 대여 사업부터 시작한 넷플릭스는 온라인 스트리밍을 위주로 서비스하고 있는 세계 최대의 엔터테인먼트 기업입니다. 넷플릭스는 어떻게 세계 최대의 기업이 됐을까요?



넷플릭스는 사용자가 선호하는 콘텐츠를 파악한 후 이를 바탕으로 유사한 콘텐츠를 사용자에게 추천해주는 시스템을 도입했습니다. 추천 시스템이란 사용자가 일일이 검색하지 않고, 사용자의 취향에 맞는 콘텐츠를 모아서 보여줌으로써 많은 사용자에게 호평받은 서비스입니다.

넷플릭스의 추천 시스템에는 머신러닝으로 만들었습니다. 머신러닝은 인공지능의 한 분야로서, 여러 가지 서비스를 만드는 데 합니다. 그럼 인공지능과 머신러닝은 무엇일까요?



☺ 인공지능이란 무엇인가?

1) '인공지능'을 인터넷으로 검색하여, 찾은 결과를 다양한 방식으로 정리해 봅시다.

2) '인공지능'을 인터넷으로 검색한 결과 중 반복되는 단어를 찾아, 몇 번 반복되었는지 정리해봅시다.

단어	횟수	단어	횟수

3) 가장 많이 나온 단어 3개를 활용하여, '인공지능'의 정의를 내려봅시다.

4) 인공지능의 실제 정의와 자신이 내린 정의와 다른 부분을 밑줄로 표시해 보고, 무엇이 다른지 적어봅시다.

☺ 머신러닝이란 무엇인가?

1) '머신러닝'을 인터넷으로 검색하여, 찾은 결과를 다양한 방식으로 정리해 봅시다.

2) '머신러닝'을 인터넷으로 검색한 결과 중 반복되는 단어를 찾아, 몇 번 반복되었는지 정리해봅시다.

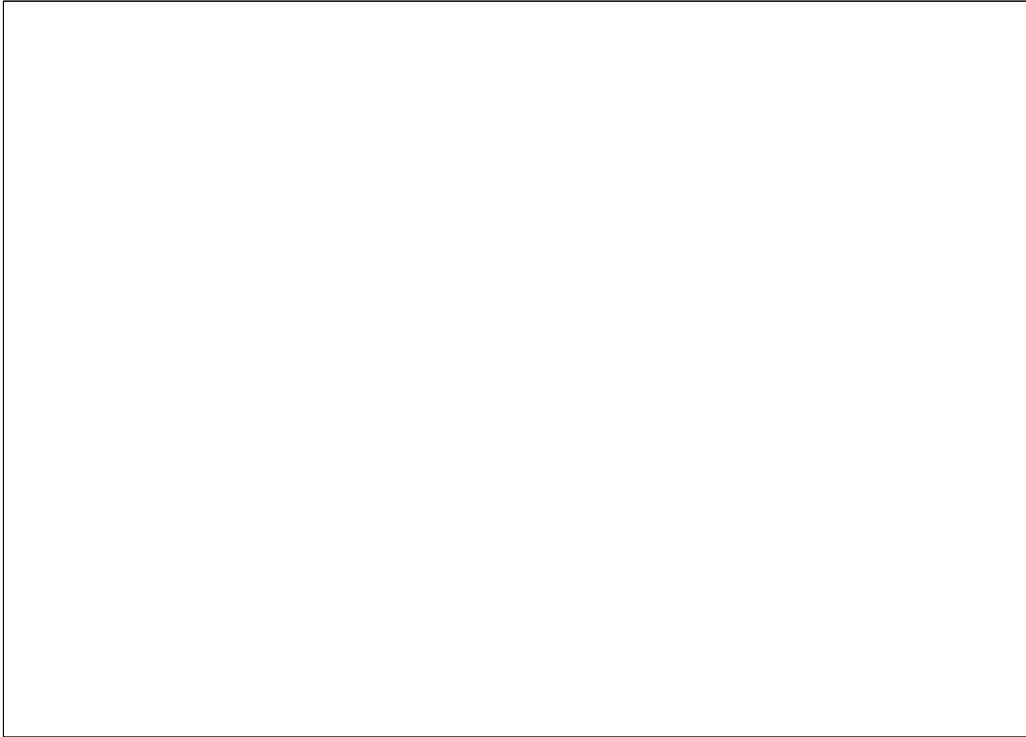
단어	횟수	단어	횟수

3) 가장 많이 나온 단어 3개를 활용하여, '머신러닝'의 정의를 내려봅시다.

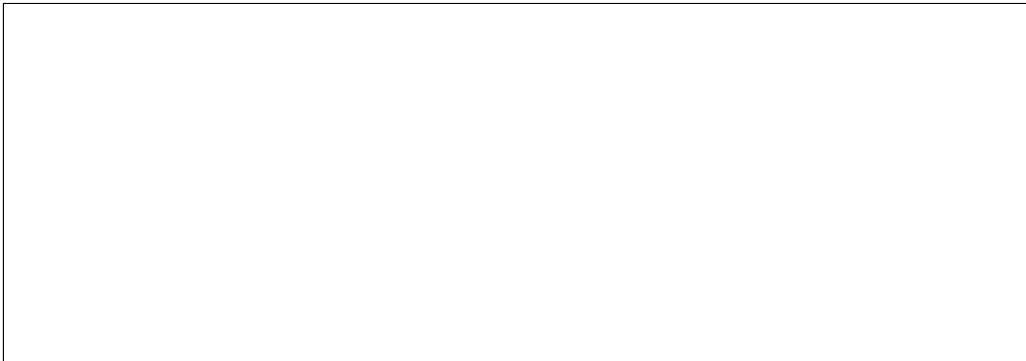
4) 머신러닝의 실제 정의와 자신이 내린 정의와 다른 부분을 밑줄로 표시해 보고, 무엇이 다른지 적어봅시다.

☺ **내가 만들고 싶은 머신러닝 서비스**

1) 인터넷을 통해 머신러닝으로 만들어진 서비스가 어떤 것이 있는지 찾아봅시다.



2) 인터넷으로 찾은 결과를 참고하여, 내가 만들고 싶은 머신러닝 서비스를 표현해봅시다.



II. 데이터 라벨링이란 무엇인가?

학습 목표 | 데이터 라벨링의 정의를 이해하고, 데이터 라벨링 과정에 관해 설명할 수 있다.

넷플릭스는 어떻게 내가 좋아하는 영화를 추천해줄까?

넷플릭스는 신작이 들어오면, 콘텐츠 팀이 해당 영화, 드라마, 애니메이션을 일일이 감상한다고 합니다. 그다음 엑셀 스프레드시트에 해당 영화와 관련 있다고 생각되는 모든 태그(꼬리표)를 입력한다고 합니다. 이때 입력하는 태그는 엄청 많이, 그리고 되도록 자세하게 입력합니다. 입력된 태그는 사용자가 가입할 때 활용됩니다.

테스트 님, 좋아하는 콘텐츠를 3개 선택하세요.

회원님이 좋아하실 만한 영화와 TV 프로그램을 더 많이 추천해 드릴 수 있습니다. 마음에 드는 콘텐츠를 선택하세요.

다음



사용자가 처음 넷플릭스에 가입하면 자신의 취향에 맞는 콘텐츠 3개를 고르게 됩니다. 그리고 3개의 콘텐츠에 붙은 태그를 바탕으로 컴퓨터 알고리즘이 사용자 취향에 맞는 콘텐츠를 찾아줍니다. 태그의 일치도가 높은 콘텐츠가 우선 노출됩니다. 이후 사용자가 넷플릭스의 콘텐츠를 많이 감상하면 감상할수록 더욱 정확한 결과가 나옵니다. 머신러닝을 바탕으로 넷플릭스의 클라우드 컴퓨팅 시스템이 수많은 태그를 일일이 대조한 후 사용자 취향에 맞는 콘텐츠를 찾아줍니다. 태그는 영어로만 입력하는 것이 아닙니다. 태그

도 현지화됩니다. 해당 국가의 문화와 언어에 맞춰서 다양한 언어로 태그를 매깁니다. 나라 별로 취향이 천차만별이기 때문입니다. 그래서 넷플릭스의 메인 화면은 사용자별로 전혀 다르다. 7,500만 명의 넷플릭스 가입자가 있으면, 7,500만 명의 넷플릭스 메인 화면이 존재합니다.

데이터 라벨링이란 넷플릭스에서 콘텐츠 팀이 영화나 애니메이션에 관련 있다고 생각하는 태그를 다는 것을 말합니다. 기계가 인간 대신 영화나 애니메이션을 추천하기 위해 필요한 과정이 데이터 라벨링입니다.

☺ **데이터 라벨링이란 무엇인가?**

1) '데이터 라벨링'을 인터넷으로 검색하여, 찾은 결과를 다양한 방식으로 정리해봅시다.

2) '데이터 라벨링'을 인터넷으로 검색한 결과 중 반복되는 단어를 찾아, 몇 번 반복되었는지 정리해봅시다.

단어	횟수	단어	횟수

3) 가장 많이 나온 단어 3개를 활용하여, '데이터 라벨링'의 정의를 내려봅시다.

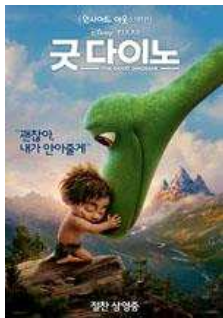
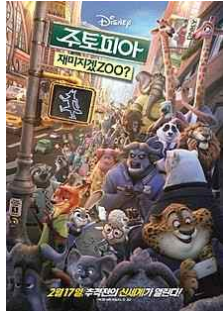
4) 데이터 라벨링의 실제 정의와 자신이 내린 정의와 다른 부분을 밑줄로 표시해보고, 무엇이 다른지 적어봅시다.

☺ **데이터 라벨링은 어떻게 할까?**

1) 인터넷에서 '애니메이션'을 검색하여, 검색 결과로 나온 단어를 적어봅시다.

2) 1) 번에서 적은 단어 중 아래 영화들을 검색했을 때 나오는 단어들을 적어봅시다.

영화 제목	단어
 <p>진짜 나를 만날 시간 Disney Pixar 인사이드 아웃 필립 실명중</p>	
 <p>Disney Pixar 업 디즈니 픽사 3D 7월 10일 개봉</p>	





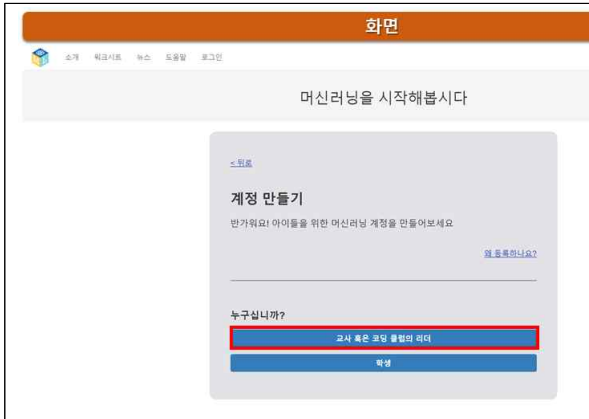
Ⅲ. '머신러닝 포 키즈' 활용하기

학습 목표 | '머신러닝 포 키즈'에 가입하여 활용하는 방법을 설명할 수 있다.

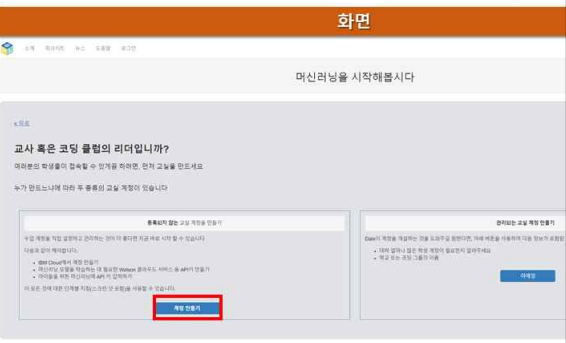
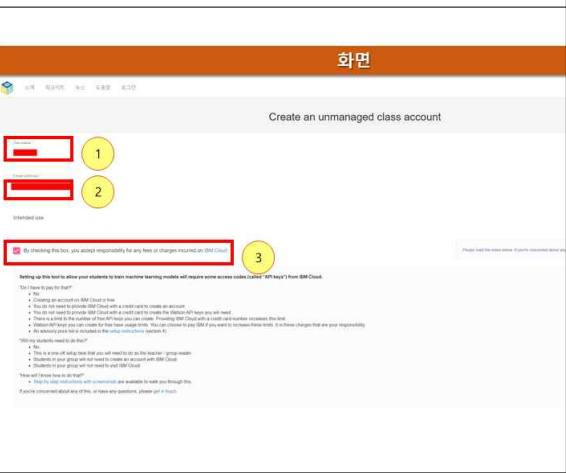
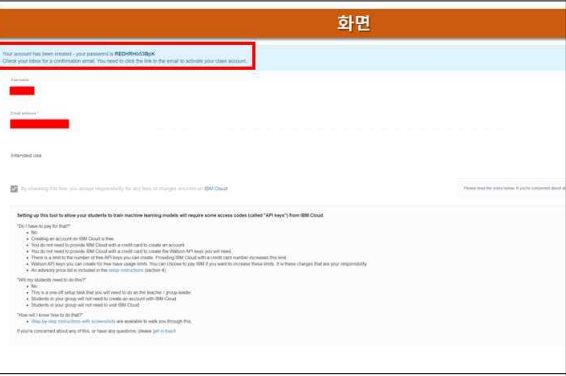

'머신러닝 포 키즈'는 인공지능을 활용한 게임을 만들 수 있는 프로그램입니다. '머신러닝 포 키즈'에 가입하여, 다양한 인공지능 게임을 만들어봅시다.

1) 계정 만들기 (자료참고 : <https://blog.naver.com/kyutto/221986592638>)

	<p>① 크롬 브라우저를 실행하여 https://machinelearningforkids.co.uk/ 에 접속한다.</p> <p>② [Language]를 클릭하고, [kr 한국어]를 선택한다.</p> <p>③ [로그인]을 클릭한다.</p>
	<p>① [계정 만들기]를 클릭한다.</p>
	<p>① [교사 혹은 코딩 클럽의 리더]를 클릭한다.</p>

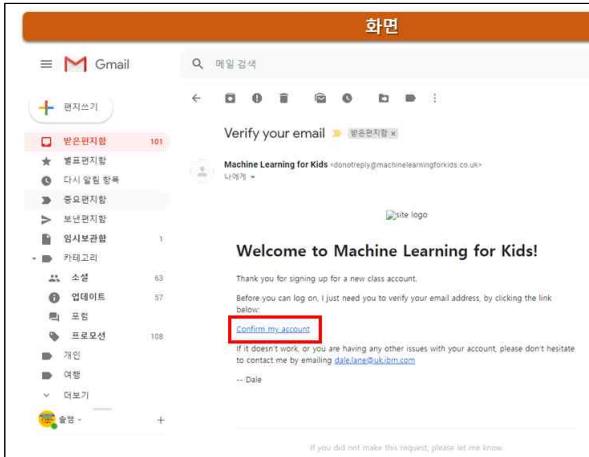


※ 학생 계정은 교사 계정을 통해 발급할 수 있으며, 꼭 교사가 아니어도 누구나 교사 계정으로 발급 가능함.

	<p>① [계정 만들기]를 클릭한다.</p>
	<p>① 사용자 이름을 입력한다. ② 이메일 주소를 입력한다. (머신러닝 포 키즈 아이디로 사용됨) ※구글 계정을 사용하는 것이 안정적이다. ③ IBM Cloud 사용을 위해 동의 체크를 한다. ④ [CREATE CLASS ACCOUNT]를 클릭한다.</p>
	<p>① 로그인을 위한 임시번호가 발급된다. 비밀번호를 변경할 것이므로 기억할 필요는 없다. ② 이전에 입력했던 이메일 주소로 회원가입 관련 인증 메일이 오니 확인하도록 한다.</p>
	<p>① 이메일 계정으로 접속하여 인증 메일을 확인하고 [Confirm my account]를 클릭한다.</p>

<p>The screenshot shows a Gmail interface with a sidebar on the left containing folders like '받은편지함' (101), '보낸편지함', '삭제된 편지함', '휴지통', '프로모션' (108), and '대보관'. The main content area displays an email titled 'Verify your email' from 'Machine Learning for Kids'. The email body says 'Welcome to Machine Learning for Kids!' and includes a link 'Confirm my account' which is highlighted with a red rectangular box. Below the link, it says 'If it doesn't work, or you are having any other issues with your account, please don't hesitate to contact me by emailing duke@mlforkids.com'.</p>	<p>※ 이메일이 오지 않았을 경우 인증 메일이 '스팸 편지함'에 있을 가능성이 크다. 스팸 편지함을 확인해보도록 한다.</p>
--	---

	<p>① [계정 만들기]를 클릭한다.</p>
	<p>① 사용자 이름을 입력한다. ② 이메일 주소를 입력한다. (머신러닝 포 키즈 아이디로 사용됨) ※구글 계정을 사용하는 것이 안정적이다. ③ IBM Cloud 사용을 위해 동의 체크를 한다. ④ [CREATE CLASS ACCOUNT]를 클릭한다.</p>
	<p>① 로그인을 위한 임시번호가 발급된다. 비밀번호를 변경할 것이므로 기억할 필요는 없다. ② 이전에 입력했던 이메일 주소로 회원가입 관련 인증 메일이 오니 확인하도록 한다.</p>
	<p>① 이메일 계정으로 접속하여 인증 메일을 확인하고 [Confirm my account]를 클릭한다. ※ 이메일이 오지 않았을 경</p>



우 인증 메일이 '스팸 편지함'에 있을 가능성이 크다. 스팸 편지함을 확인해보도록 한다.

IV. 신문 헤드라인 정복하기

학습 목표 | 신문의 헤드라인을 보고, 어떤 신문의 헤드라인인지 맞출 수 있는 머신러닝을 만드는 방법을 설명할 수 있다.

신문 기사들의 첫 제목(헤드라인)을 보면 신문사마다 강조하는 부분들이 다릅니다. 인공지능에 서로 다른 성격들을 지닌 문장을 알아보고 어떤 신문사에서 만든 기사인지 분류할 수 있게 학습시켜봅시다.

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝 포 키즈에서 '워크시트' 메뉴에서 선택합니다.)

[학습데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터 종류	
데이터 라벨	
라벨별로 데이터 입력 기준 만들기 (예시 포함)	
최소로 입력할 데이터 수	

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝 포 키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도

V. 트위터는 어떻게 생각하나요?

학습 목표 | 트위터에 검색한 주제를 스스로 정한 기준에 맞춰 분류하는 머신러닝을 만드는 방법을 설명할 수 있다.

세상에는 많은 사람이 있고 그 사람들은 어떠한 주제에 대하여 각각 다른 생각하고 있습니다. 내가 정한 주제에 대하여 사람들은 어떤 생각하고 있으며 이를 트위터에 어떻게 표현하고 있을까요?

☺ 머신러닝 모델 만들기

[만들고 싶은 주제 정하기] 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.

[모델 정하기] 우리 팀이 정한 주제에 맞는 머신러닝 모델을 정해봅시다.
(머신러닝 포 키즈에서 '워크시트' 메뉴에서 선택합니다.)

[학습데이터 기준 정하기] 주제에 맞는 데이터를 아래 표에 정리해봅시다.

데이터 종류	
데이터 라벨	
라벨별로 데이터 입력 기준 만들기 (예시 포함)	
최소로 입력할 데이터 수	

[데이터 입력하기] 위의 표로 정리한 기준에 따라 '머신러닝 포 키즈' 프로젝트에 데이터를 입력합니다.

[훈련된 모델 정확도 테스트하기] 훈련된 모델에 다른 데이터를 넣어, 만든 모델의 정확도를 적어봅시다. 정확도가 80% 이하일 경우, 데이터를 추가로 넣어 정확도를 높여 봅시다.

1차 정확도	2차 정확도	3차 정확도	4차 정확도	5차 정확도