

데이터 라벨링 중심의 머신러닝 교육이 초등학생 컴퓨팅 사고력에 미치는 효과

문우종* · 김봄솔* · 김정아* · 김봉철* · 서영호* · 오정철** · 김용민*** · 김종훈*
제주대학교*, 도평초등학교**, 제주시교육지원청***

요약

본 연구는 초등학생의 컴퓨팅 사고력을 향상시키기 위한 교육 방법으로 데이터 라벨링 중심의 머신러닝 교육 프로그램을 개발하여 적용한 후 그 효과를 검증하였다. 교육 프로그램은 현직 초등학교 교사 100명을 대상으로 실시한 사전 요구분석 결과를 바탕으로 설계 및 개발을 진행하였다. 개발한 교육 프로그램의 효과를 검증하기 위하여 K 초등학교에 재학 중인 6학년 학생 17명을 대상으로 1일 2차시씩 총 6주간 12차시의 교육을 진행하였다. 해당 교육이 컴퓨팅 사고력 향상에 미친 효과를 측정하기 위해 '비버챌린지(Bebras Challenge)'를 활용하여 사전 사후 검사를 진행하여 교육적 효과를 분석하였다. 분석 결과 데이터 라벨링 중심의 머신러닝 교육이 초등학생의 컴퓨팅 사고력 향상에 기여한 것으로 나타났다.

키워드 : 데이터 라벨링, 인공지능, 머신러닝 교육, 머신러닝 포 키즈, 컴퓨팅 사고력

Effect of Machine Learning Education Focused on Data Labeling on Computational Thinking of Elementary School Students

Woojong Moon* · Bomsol Kim* · Jungah Kim* · Bongchul Kim* · Youngho Seo*
Jeongcheol OH** · Yongmin Kim*** · Jonghoon Kim*

Jeju National University* · Dopyeong Elementary School** · Jeju City Office of Education***

Abstract

This study verified the effectiveness of machine learning education programs focused on data labeling as an educational method for improving computational thinking of elementary school students. The education program was designed and developed based on the results of a preliminary demand analysis conducted on 100 elementary school teachers. In order to verify the effectiveness of the developed education program, 17 sixth-grade students attending K Elementary School were given 2 classes per day for a total of 6 weeks. In order to measure the effect of the training on improving computational thinking, the educational effects were analyzed by conducting pre-post-inspection using the "Beaver Challenge". According to the analysis, machine learning education focused on data labeling contributed to improving computational thinking of elementary school students.

Keywords : Data Labeling, AI, Machine Learning, Machine Learning for Kids, Computational Thinking

본 논문은 2021년도 제주대학교 교육·연구 및 학생지도비 지원에 의하여 연구되었음.

교신저자 : 김종훈(제주대학교 초등컴퓨터교육전공)

논문투고 : 2020-12-03

논문심사 : 2020-12-03

심사완료 : 2020-12-28

1. 서론

4차 산업 혁명 시대가 가속화되면서, 4차 산업 혁명을 이끄는 IT 기술 교육의 중요성이 커지고 있다. 특히 인공지능(Artificial Intelligence: AI)의 발전이 가속화되면서, 산업과 사회, 인간의 삶 전반에 걸친 패러다임이 변화하고 있다. 이런 시류에 맞춰 한국정보과학교육연합회에서는 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. 2019년 한국창의과학재단에서 발표한 차세대 소프트웨어 교육 표준모델에서는 인공지능 개념, 지식표현 및 추론, 머신러닝, 인공지능경망에 관한 교육을 초등학교 3학년부터 고등학교 2학년을 대상으로 시행할 것을 제시하였으나 현재 적용되지 않고 있으며, 소프트웨어 교육과 융합하는 방향으로 ‘인공지능과 융합’ 영역이 2019년에 도입되었다[15].

본 연구에서는 이러한 교육 현장의 상황을 고려하여 초등학교 6학년을 대상으로 12차시 분량의 인공지능과 머신러닝의 개념을 학습하고, 나아가 머신러닝의 성능에 영향을 미치는 데이터 라벨링의 개념을 익혀 데이터를 수집, 분류, 정제하는 과정을 거쳐 머신러닝 모델을 학습시키고, 학습시킨 모델을 활용하여 블록 코딩으로 인공지능 프로그램을 만들어 보는 교육 프로그램을 개발하여 적용하였다. 개발한 교육 프로그램은 사전 요구분석 결과를 분석하고 교육 도구와 세부 주제를 선정하였다[4].

2. 이론적 배경

2.1. 머신러닝

머신러닝은 학습 데이터의 패턴과 규칙을 컴퓨터가 스스로 추출하는 기술로서, 인공지능에의 ‘전문가 시스템’의 한계점이었던 ‘유연한 대응’을 극복한 방식이다. 현재 인공지능 서비스라고 소개되는 시스템은 머신러닝을 기반으로 개발된 방식이며, 인공지능으로 구현했다는

표현보다는 머신러닝으로 구현했다는 표현이 좀 더 정확한 표현이다[7].

머신러닝은 컴퓨터가 학습하는 방식에 따라, 지도학습, 비지도 학습이 있으며, 강화 학습으로 분류한다. 지도학습은 정답이 있는 데이터를 활용하여 학습을 진행하며, 훈련 데이터를 학습한 결과를 바탕으로 테스트 데이터의 정답을 맞히는 방식으로 학습 모델이 구성된다. 비지도 학습은 정답이 주어지지 않는 데이터를 비슷한 유형으로 나눠 군집화를 하며, 기계가 스스로 데이터의 특성을 파악하는 학습 방식이다[3]. 강화 학습은 특정 조건에서 성공했을 때 일정한 점수를 주어 학습을 진행하는 방식이다.

2.2. 머신러닝 포 키즈

‘머신러닝 포 키즈’는 IBM에서 제공하는 머신러닝 교육 서비스이다[13]. IBM에서 개발한 머신러닝 시스템인 왓슨(Watson)을 활용하여, 숫자, 텍스트, 이미지, 소리 데이터를 바탕으로 머신러닝을 진행할 수 있다[11]. 2020년 7월 31일 현재 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 프로젝트를 선택한 뒤 실습을 진행할 수 있으며, 선택한 프로젝트의 모델을 응용하여 변형된 프로젝트를 만들 수 있다.

2.3. 데이터 라벨링(레이블링)

데이터 라벨링은 머신러닝의 학습에 사용할 데이터 샘플을 검출하고 태그(라벨)를 붙이는 과정을 말한다. 데이터 라벨링은 머신러닝의 학습을 위한 중요한 사전 작업이며, 특히 지도학습의 경우 입력력 데이터에 모든 분류 라벨을 부착하여, 머신러닝 모델이 학습을 진행할 때 정답 여부의 근거를 제공한다[15]. 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.

이보배(2020)는 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[8].

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

2.4. 컴퓨팅 사고력

윙(Wing)이 주창한 컴퓨팅 사고력은 소프트웨어 교육 시 교수자가 학습 전략을 수립할 때 고려해야 하는 핵심 역량으로 평가받고 있다[17]. 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[18]. ISTE의 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[2].

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. 이영준(2014)은 초·중등 단계에서 컴퓨팅 사고력 도입을 위한 기초연구에서 초·중등 교육과정에 CT를 도입하기 위해 교과별 CT 교육 프로그램을 개발하여 제시하였다[12].

머신러닝 교육 또한 소프트웨어 교육의 한 분야로서 학습자의 컴퓨팅 사고력을 향상하는 방향으로 학습 전략을 수립하여 교육 프로그램을 개발하려는 다양한 연구가 진행 중이다.

2.5. 선행연구 분석

김갑수(2017)는 초등학생의 인공지능 교육을 위한 교수 학습 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 제안했다[5].

이은경(2020)은 국내외 초·중등학교 인공지능 교육과정 분석에서 인공지능 교육과정 개발을 위한 방향 및 단계를 구체적으로 설정할 필요가 있음을 제안한다. 미국이나 EU의 경우 별도의 인공지능 교육을 위한 교육과정이나 내용을 제시하고 있으나, 한국은 차세대 소프트웨어 교육과정 표준 모델에 인공지능 관련 영역을 포함함으로써 AI 교육을 위해 필수적인 교육 요소들을 포함하지 못하는 한계를 지니고 있다고 제안했다[9].

이영호(2019)는 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[10].

본 연구에서는 이러한 선행연구의 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.

3. 데이터 라벨링 중심의 머신러닝 교육

3.1. 연구 절차

본 연구에서는 Dick&Carey의 ADDIE 모형의 절차에 따라 교수 체제 설계 과정의 일반적인 형태를 고려하여 교육 프로그램을 <Table 1>과 같이 개발하였다.

<Table 1> The Theme of education program

Analysis	Demand Analysis (for elementary school teachers)
Design	· Specification of performance objectives - Effect of Machine Learning Education Program Using Data Labeling · Evaluation tool design
Development	· Development of Teaching Materials - 12nd-hour plan - Teaching Materials
Implementation	· Machine Learning Education Program Using Data Labeling
Evaluation	· Post-Evaluation with Korea Bebras Challenge and Performance check

3.2. 교육 프로그램 설계 및 개발

3.2.1. 요구분석

ADDIE 모형의 절차에 따른 요구분석은 제주도 내 초등교사 100명을 대상으로 한 설문 조사 결과를 바탕으로 실시하였다.

<Table 2> The Theme of education program

well-informed	be average	not well aware
32(32%)	33(33%)	35(35%)

머신러닝(인공지능)에 대해 얼마나 알고 있는지 조사한 결과는 <Table 2>와 같다. 초등교사 중 32%가 머신러닝(인공지능)에 대해 잘 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.

<Table 3> The need of Machine Learning (AI) education

well-informed	be average	not well aware
51(78%)	13(20%)	1(2%)

머신러닝(인공지능) 교육이 필요하다고 생각하는지 조사한 결과는 <Table 3>과 같다. 초등교사 중 78%는 머신러닝(인공지능) 교육이 꼭 필요하다고 응답했고, 필요하지 않다고 응답한 경우는 2%에 그쳤다.

<Table 4> Understanding Data Labeling

well-informed	be average	not well aware
37(57%)	21(32%)	7(11%)

데이터 라벨링(레이블링)에 대해 얼마나 알고 있는지 조사한 결과는 <Table 4>와 같다. 초등교사 중 55%가 데이터 라벨링에 대해 잘 알고 있다고 응답했으며, 전혀 알지 못한다고 응답한 경우는 13%로 나타났다.

<Table 5> The need of Data Labeling education

well-informed	be average	not well aware
36(55%)	21(32%)	8(13%)

데이터 라벨링 교육이 필요하다고 생각하는지 조사한 결과는 <Table 5>와 같다. 초등교사 중 57%는 데이터 라벨링 교육이 꼭 필요하다고 응답했고, 필요하지 않다고 응답한 경우는 11%에 그쳤다.

사전 요구분석 결과 머신러닝에 대해 알고 있다고 응답한 교사들은 데이터 라벨링(레이블링)의 개념 역시 이해하고 있으며, 데이터 라벨링 교육의 필요성을 공감하고 있었다. 데이터 라벨링은 머신러닝 학습을 위한 중요한 사전 작업으로서 데이터 라벨링을 어떻게 하느냐에 따라 수행 결과가 달라지기도 한다. 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.

3.2.2. 프로그램 설계

사전 요구분석과 학습자의 수준을 고려하여 전체 12차시로 구성된 데이터 라벨링 중심의 머신러닝 교육 프로그램을 <Table 6>과 같이 설계하였다.

<Table 6> The Theme of education program

Hour	Learning theme
1~3	- Introduction to Machine Learning - Preliminary-test, Data labeling foundation
4~9	- Learning the Concept of Machine - Learning Using Machine Learning for Kids
10~12	- Team project work - Post-test

머신러닝에 대한 개념이 초등학생들에게 어렵다는 요구분석의 응답 결과를 고려하여, 1~3차시는 인공지능의 기본 개념을 익히는 활동지를 활용한 언플러그드 활동을 통해 학습자가 직접 찾아보고, 관련 자료를 정리하면서 능동적으로 개념을 익힐 수 있도록 구성하였다. 4~9차시는 머신러닝 포 키즈를 활용한 블록형 프로그래밍 언어 기반의 교육 프로그램을 조별 활동으로 데이터를 수집, 분류 및 정제하는 과정을 실습하며, 머신러닝의 구현 원리를 실습해볼 수 있게 구성하였다. 10~12차시는 4~9차시에서 배운 내용을 바탕으로 자율주제를 선정하여 조별 활동을 진행하도록 구성하였다. 이를 통해 적극적으로 학습에 참여하도록 지도하였다.

3.2.3. 프로그램 개발

프로그램 설계를 토대로 개발한 전체 12차시의 학습 내용은 <Table 7>과 같다.

<Table 7> The Theme of education program

Lesson	Topic	Objectives
1st	What is artificial intelligence?	1.Explore the concept of AI 2. Experience various AI services.
2nd	What is data labeling?	1. Explain what data labeling is about. 2. A method of labeling data can be performed.

Lesson	Topic	Objectives
3rd	Using 'Machine Learning for Kids'	Types of machine learning and how they are used in 'Machine Learning for Kids'
4th	Newspaper headlines 1 -Planning	Design the process of making machine learning to classify newspaper headlines
5th	Newspaper headlines 2 -Data labeling	Collect newspaper headlines and perform data labeling
6th	Newspaper headlines 3 - Create a program	Program coding to classify headlines using learned machine learning
7th	What do you think of Twitter 1 -Planning	Learn how to create machine learning to classify Twitter timeline
8th	What do you think of Twitter 2 -Data labeling	Learn machine learning by collecting Twitter timeline and conducting data labeling
9th	What do you think of Twitter 3 -Create a program	Coding a program that classifies Twitter's timeline using machine learning
10th	Machine learning we made 1 -Planning	Plan to make machine learning with assorted activities
11th	Machine learning we made 2 -Data labeling	Collect and classify data in assorted activities
12th	Machine learning we made 3 -Create a program	Using the learned machine learning, program coding according to the prepared plan

1차시~3차시까지 진행한 언플러그드 활동을 통해 인공지능 서비스를 소개하고, 해당 서비스에서 데이터 라벨링을 진행하는 방식과 학습된 머신러닝 모델을 어떻게 활용하는지 전달하였다.

4차~9차시는 머신러닝 포 키즈에 제시된 프로젝트를 활용하되 활동지를 통해 데이터 라벨링을 기획하는 단계를 실습할 수 있게 구성하였다. 학습자가 데이터를 수집 하는 방법에 대해 학습자가 조별 활동으로 정하여 데이터를 직접 수집 및 정리하여, 데이터 라벨링하는 과정을 구성하였다. 이는 선행 연구된 인공지능 교육 프로그램에서 기존의 검증된 데이터를 활용하는 것과 차별화된 단계로, 학습자가 데이터 라벨링을 직접 체험해보도록 구성하였다. 또한 훈련한 모델을 평가하는 활동을 넣어, 모델의 정확도를 높이기 위해 데이터 라벨링 과정의 성찰하고, 재학습 시키는 교육활동을 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 만들어 보는 차시로 구성하였다.

4. 연구 방법 및 절차

4.1. 연구 가설

연구가설 : 데이터 라벨링 중심의 머신러닝 교육에 의한 학습자의 사전, 사후 컴퓨팅 사고력은 차이가 있다.

대립가설 : 데이터 라벨링 중심의 머신러닝 교육에 의한 학습자의 사전, 사후 컴퓨팅 사고력은 차이가 없다.

4.2. 연구 대상

본 연구는 제주시 K 초등학교에 재학 중인 6학년 학생 17명을 대상으로 진행하였다. 참여한 학생들은 스크래치와 엔트리를 활용한 소프트웨어 교육을 받은 경험이 있지만, 인공지능에 대한 교육 경험이 없었다. 연구 대상의 성별은 <Table 8>과 같다

<Table 8> gender of the subject

Male	Female	Total
10	7	17

4.3. 연구 도구

본 연구에서는 컴퓨팅 사고력을 측정하는 도구로 비버챌린지(Bebras Challenge)(2018) 한국 그룹 III의 도전하기 문항을 사용하였다. 학생들의 컴퓨팅 사고력을 평가하기 위해 개발된 비버챌린지는 과제 해결 중심의 평가 모델이다[1]. 김은지는 컴퓨팅 사고력 평가도구로써 비버챌린지 2017 신뢰도 분석 : 한국 그룹 III를 중심으로(2018)에서 비버챌린지 문항의 신뢰도가 Cronbach's Alpha = .784로써 지능검사 같은 인지적 검사에서 0.7을

쓸 수 있다는 Kline의 주장을 소개하며, 문항의 신뢰도가 높다고 평가하였다. 이는 비버챌린지가 컴퓨팅 사고력을 일관성 있게 측정할 수 있다는 것을 의미한다[6]. 본 연구에서는 박은주의 알고리즘 교육을 통한 비버챌린지 결과 분석(2019)에서 제시한 문항별 난이도를 반영하여[16], 비버챌린지에서 제시하는 채점 기준에 맞춰 채점하였으며[1], 검사 도구로 비버챌린지를 사용하여 사전-사후검사 통제집단설계(pre test-post test control group design)를 바탕으로 연구를 진행하였다.

4.4. 연구 설계 및 처치

데이터 라벨링 중심의 머신러닝 교육을 실시하기 전에 비버챌린지(Bebras Challenge) 사전 검사를 실시하였다. 교육 프로그램은 6주간 총 12차시의 내용으로 구성되어, 진행하였다. 교육 종료 후 실험집단의 교육 효과를 검증하기 위해 비버챌린지(Bebras Challenge) 사후 검사를 실시하였으며, 본 연구의 설계를 도식화하여, <Table 9>에 제시하였다.

<Table 9> Experimental design

	Pre-test	Treatment	Post-test
G	O1	X	O2

G: Experimental group(N = 17)
 O1, O2: Bebras Challenge(Figure A, B style)
 X: Machine learning education program using data labeling

5. 연구 결과

5.1. 교육 프로그램 효과 검증

데이터 라벨링 중심의 머신러닝 교육이 컴퓨팅 사고력 향상에 영향을 주었는지 분석하기 위하여 검증을 진행하였다.

5.1.1. 컴퓨팅 사고력 검사 정규성 검증

실험집단의 표본의 크기가 17명으로 '10<n<30'에 위치하기 때문에, 실험집단의 컴퓨팅 사고력의 사전 검사와 사후 검사 결과가 정규성을 만족하는지 확인하기 위해 정규성 검정을 실시하였다. 정규성 검정은 보편적으로 활용하는 샤피로-윌크(Shapiro-Wilks) 검사를 활용하였으며, 해당 결과는 <Table 10>과 같다.

<Table 10> Normality test

Subscales	Descriptive Statistics(N=17)				stat	p
	M	SD	Max	Min		
Pre-test	18.000	22.226	66.00	-18.00	.955	.538
Post-test	32.352	22.363	82.00	-6.00	.963	.690

*p<.05

정규성 검정 결과 사전 검사 결과와 사후 검사 결과가 모두 정규성을 만족한 것으로 나타났다.

5.1.2. 컴퓨팅 사고력 사전-사후 검사 비교

사전-사후 검사 결과 샤피로-윌크(Shapiro-Wilks) 검사에서 정규성을 확보하여, 대응표본 t검정을 시행하였다. 대응표본 t검정 결과는 <Table 11>에 제시하였다.

<Table 11> Computational Thinking(Paired T-test)

Subscales	N	Pre-Test		Post-Test		t	p
		M	SD	M	SD		
CT	17	18.000	22.226	32.352	22.363	-3.727	.002*

*p<.05

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

5.1.3. 학생 산출물 심층 분석

본 연구에서 개발한 데이터 라벨링 중심의 머신러닝 교육 프로그램은 K 초등학교 6학년을 대상으로 6주에 걸쳐 총 12차시에 걸쳐 진행되었다. 학생들의 4~9차시의 산출물과 10~12차시의 산출물을 비교는 <Table 12>와 같다. <Table 12>를 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 학습에 필요한 단어 수를 최소 10개에서 최대 20개까지 늘어나 학습하는 데이터의 개수가 많을수록 정확한 결과가 나온다는 것을 체득하였음을 알 수 있었다.

<Table 12> Comparison of students' output

4th~9th Activity Paper	10th~12th Activity Paper
<p>① 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>② 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>③ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>④ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>⑤ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>⑥ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p>	<p>① 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>② 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>③ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>④ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>⑤ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p> <p>⑥ 머신러닝 모델 만들기 (단종과 같은 주제 정하기 우리 팀이 만들고 싶은 머신러닝 주제를 적어봅시다.)</p>

5.2. 연구 결과 분석

비버챌린지 결과를 분석하기 전 실험집단이 정규분포를 갖췄는지 확인하기 위해 샤피로-윌크 검사를 시행하였다. 정규성 검정 결과 정규성을 확보하여, 대응표본 t검정을 진행하였다. 대응표본 t검정 결과 컴퓨팅 사고력은 사전 검사 대비하여 평균 점수도 상승하였으며, 유의확률도 통계적으로 유의미한 향상을 보였다. 또한 학생들의 산출물을 심층 분석한 결과 데이터를 수집할 때 기준이 되는 단어나 선정 기준을 고민하여 전략적으로 선정하는 등 데이터 라벨링에 중심을 둔 교육을 통해 컴퓨팅 사고력 향상에 긍정적인 영향을 미치는 것으로 보인다.

따라서 대립가설을 기각하고 연구가설을 채택하여 본 교육에 의한 학습자의 사전, 사후 컴퓨팅 사고력은 차이가 있으며, 컴퓨팅 사고력 향상에 긍정적임을 제시하였다.

6. 결론

본 연구는 데이터 라벨링 중심의 머신러닝 교육이 초등학교 학생들의 컴퓨팅 사고력에 미치는 영향을 파악하고자 하였다. 이를 위해 ADDIE 모형의 절차에 따라 사전 요구분석 결과를 바탕으로 교육 프로그램을 개발하였고, 6주에 걸쳐 총 12시간 동안 데이터 라벨링에 중점을 둔 머신러닝 교육 프로그램을 운영하였다. 교육 전후에 사전-사후 검사를 통해 컴퓨팅 사고력의 변화를 검증하였다.

검사 결과 본 연구에서 개발한 교육 프로그램은 초등 학생의 컴퓨팅 사고력 향상에 효과적인 것으로 나타났다. 학생들은 효과적인 머신러닝 프로그램을 구현하기 위해 데이터 라벨링이 어떤 역할을 하는지 알아보고, 라벨별로 데이터의 기준을 만들며 머신러닝에 학습에 대한 만족도도 높았다.

다만, 본 연구의 실험집단은 일반적인 상관관계 연구에 필요한 30명보다 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 없다는 문제가 있다.

후후의 연구에서는 다수의 참가자를 확보하고, 실험 집단과 비교집단을 구성하여 연구 결과에 대한 각 요인들을 좀 더 체계적으로 분석할 필요가 있다.

참고문헌

- [1] Bebras Challenge(2020). Introduction to Bebras Challenge. <https://bebraschallenge.org/>
- [2] ISTE Introduction to ISTE. <https://www.iste.org/>
- [3] Jang. J.H.(2018). New Artificial Intelligence Technology Self-study Artificial Intelligence. Introduction to SAMSUNG SDS. <https://www.samsungsds.com/global/ko/support/insights/Generative-adversarial-network-AI.html>
- [4] Kim. B.S, Moon. W.J.(2020). Development of Machine Learning Education Program Using Data Labeling. *Proceeding of the KAIE Summer Conference 2020*, 157-163
- [5] Kim. K.S, Park Y K(2017). A Development and Application of the Teaching and Learning Model of Artificial Intelligence Education for Elementary Students. *Journal of The Korean Association of information Education*, 21(1), 137-147.
- [6] Kim E.J(2018). Beaver Challenge 2017 Reliability Analysis as a Computing Thinking Assessment Tool : The Korean Group III
- [7] Kanako Onishi(2019). *Easiest AI Introduction*, 1st ED. Seoul: Atio.
- [8] Lee. B.B.(2020). Impact of Data Imbalance on Machine Learning Model Performance.
- [9] Lee. E.K.(2020). A Comparative Analysis of Contents Related to Artificial Intelligence in National and International K-12 Curriculum. *Journal of The Korean Association of information Education*, 23(1), 37-44.
- [10] Lee. Y.H.(2019). An Analysis of the Influence of Block-type Programming Language-Based Artificial Intelligence Education on the Learner's Attitude in Artificial Intelligence. *Journal of The Korean Association of information Education*, 23(2), 189-196.
- [11] Lee. Y.H.(2020). Artificial Intelligence for Everyone.
- [12] Lee. Y.J.(2014). Research for Introducing Computational Thinking into Primary and Secondary Education.
- [13] Machine Learning For Kids(2020). Introduction to Machine Learning For Kids.
- [14] Margaret Rouse(2019). Definition of Data labeling. <https://whatis.techtarget.com/definition/data-labeling>
- [15] Oh. P.R. (2020). Introducing 'AI and convergence' in SW education. Introduction to Chosun Edu. http://edu.chosun.com/site/data/html_dir/2020/
- [16] Park.S.J.(2019). Analysis of Bebra Challenge Results through Algorithm Education.
- [17] Seo. J.Y.(2019). SW education plan for fostering next-generation SW talents based on AI education.
- [18] Wing J.M(2006). Computational Thinking.

저자소개



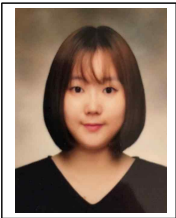
문 우 종

2019~ 제주대학교 컴퓨터교육전공
(박사과정)
현재 영평초등학교 교사
관심분야: SW교육, 프로그래밍
e-mail: mwj1006@korea.kr



서 영 호

2014 한국교원대학교 교육공학과
(교육학석사)
2017 제주대학교 컴퓨터교육전공
(박사과정 수료)
관심분야: SW교육, 디자인씽킹
e-mail: ho2832@naver.com



김 봄 솔

2019~ 제주대학교 초등컴퓨터교육
전공 (석사과정)
2019~ 현재 (주)제주넷 기업부설연
구소 책임연구원
관심분야: SW교육, 인공지능
e-mail: bomsolkim@jejunu.ac.kr



오 정 철

2020 제주대학교 컴퓨터교육전공
(교육학박사)
현재 도평초등학교 교사
관심분야: SW교육, IT피플
e-mail: love1748@korea.kr



김 정 아

2012 제주대학교 컴퓨터교육전공
(박사과정 수료)
2019~ 현재 김녕초등학교 교사
관심분야: SW교육, 언플러그드교육
e-mail: vadang@korea.kr



김 용 민

2018 제주대학교 컴퓨터교육전공
(교육학박사)
2019~ 제주시교육지원청장학사
관심분야: 데이터과학교육, AI교육
e-mail: mega11@korea.kr



김 봉 철

2021~ 제주대학교 컴퓨터교육전공
(박사과정)
현재 제주대학교교육대학부설
초등학교 교사
관심분야: 피지컬 컴퓨팅
e-mail: pankun@korea.kr



김 종 훈

2019~ 현재 제주대학교 교수
관심분야: 컴퓨터 교육
e-mail: jkim0858@jejunu.ac.kr