

공공데이터 활용 데이터사이언스 교육을 위한 초등학교 교수학습모형 개발 연구

신승기

서울교육대학교 컴퓨터교육과

요약

본 연구에서는 초등학교에서 학습자의 인지적 발달단계를 고려한 데이터사이언스 교육의 교수학습모형을 개발 하는데 목표를 두고 있으며, 공공데이터를 활용하여 데이터를 수집하고 문제해결에 필요한 데이터를 찾기 위한 전략과 과정을 제시하고자 하였다. 데이터사이언스 교육을 위한 교수학습모형은 인공지능 교육을 위한 인지적 학습 환경의 프레임워크인 Agency(인지적학습보조), Abstraction(추상화), Modeling(알고리즘구현)을 기반으로 데이터사이언스 생명주기에서 도출된 컴퓨팅사고력과 연계된 문제해결의 과정을 토대로 교수학습모형이 구성되었다. 교수 학습과정의 모든 단계에서 사회적 영향을 고려하며, 데이터 식별을 통한 수집과 발견이 강화되었다. 특히, 일상에서의 다양한 문제를 해결하기 위해 문제를 발견하고 이를 해결하기 위한 데이터를 수집하며 문제해결의 과정을 수행하면서 데이터 분석결과를 시각화하는 활동으로 교수학습모형이 구성되었다.

키워드 : 데이터사이언스, 교수학습모형, 인지적 학습환경, 공공데이터, 컴퓨팅사고력

A Study on the Instructional Model in Elementary School for Data Science Education using Public Data

Seungki Shin

Department of Computer Education, Seoul National University of Education

Abstract

This study aims to develop an instructional model of data science education that considers learners' cognitive development stages in elementary schools and to present strategies and processes to collect the data for problem-solving using public data. The instructional model for data science education is based on the problem-solving process linked to computational thinking derived from the life cycle of data science based on Agency, Abstraction, and Modeling, the framework of cognitive learning environment for artificial intelligence education. Social influences are considered at all stages of the instructional process, and collection and discovery through data identification have been strengthened. In particular, the instructional model was composed of activities to discover problems, collect data to solve problems and visualize data analysis results while performing the problem-solving process.

Keywords : Data Science, Instructional Model, Cognitive Learning Environment, Public Data, Computational Thinking

본 연구는 서울교육대학교 교내연구비 지원을 통해 수행된 연구임.

논문투고 : 2023-01-08

논문심사 : 2023-01-13

심사완료 : 2023-01-20

1. 서론

2022년은 컴퓨터교육에서 새로운 전환점으로서 큰 의미를 갖는 해이다. 우리나라에 컴퓨터교육이 학과로 설치된 연도가 1992년이라는 점을 감안하면[22] 2022년은 컴퓨터교육이 학문으로 편제되어 운영된 지 30주년을 의미한다. 학문을 세대(世代)로 구분한다면 1992년부터 2022년까지의 시기를 1세대라고 분리할 수 있으며, 새로운 30년이 시작되는 2023년부터는 2세대로 설명할 수 있겠다. 세대(世代)에서 세(世)는 열 십(十)이라는 글자가 3개가 모여 만들어짐을 생각해본다면 지나간 30년과 새로운 30년은 새로운 세대로 구분되는 시점이라고 할 수 있다.

교육정책에 대한 변화로서는 2022년에 디지털 인재 양성 종합방안을 발표함으로써 디지털 대전환 시대의 사회적인 변화에서 교육을 통한 인재 양성 및 국가경쟁력을 확보하기 위한 전략을 제시하였다[14]. 특히, 2026년까지 100만명의 디지털 인재를 양성하기 위해 전 국민의 디지털 교육 기회 확대 및 역량 강화를 지원하기 위하여 초·중등 정보교육을 확대하기 위한 전략과 방향을 발표하였다[14]. 아울러 학생들의 디지털 역량을 강화하고 모든 교과수업에서 이를 활용하는 교육환경과 내용에서의 디지털 전환의 내용을 제시하였다[14].

내용으로서의 교육의 방향에 대한 변화는 2022년에 발표된 국가수준 교육과정을 통해 살펴볼 수 있다. 언어 소양과 수리 소양 중심의 역량 강화에 대한 기존의 전통적인 교육의 범주에서 디지털 소양을 핵심 소양으로 추가함으로써 정보교육을 통한 디지털 소양 강화뿐만 아니라 모든 교과영역에서 디지털 기반의 내용체계 변화를 제시하고 있다[12]. 정보교육을 체계적으로 실시하기 위하여 컴퓨팅사고력을 기반으로 문제해결력을 기르는 과정을 수행하도록 제시하고 있으며, SW교육과 AI 교육을 통해 컴퓨팅사고력을 기를 수 있도록 구성되어 있다[13]. 2015 개정 교육과정으로부터 도입된 SW교육은 프로그래밍언어를 활용하여 피지컬컴퓨팅과 연계하여 문제해결의 알고리즘을 설계하는 것을 중점으로 다루고 있다[8][10]. 반면 AI교육은 기존의 인공지능 알고리즘을 기반으로 데이터를 입력하여 미래를 예측하거나 입력한 데이터를 분류하는 과정에 목표를 두고 있다[8]. AI교육에서 다루고자 하는 컴퓨팅사고력 기반의 문제해

결력을 기르기 위해 일상에서의 문제를 발견하고 이를 해결하기 위해 데이터를 분류하거나 예측하는 과정을 수행하게 되며, 문제해결에 필요한 적절한 데이터를 실생활에서 직접 수집하거나 이미 수집된 데이터를 활용하여 교육활동이 운영되고 있다. 데이터를 수집하고 분석하여 시각화하며 도출된 인사이트를 토대로 수행되는 교수학습의 과정을 데이터사이언스 교육이라고 할 수 있다. 즉, 데이터를 기반으로 문제해결의 아이디어를 도출하고 객관적이며 타당한 데이터 분석의 결과를 기반으로 의사결정을 수행하는 과정이라고 할 수 있다. 학교 현장에서는 데이터사이언스 교육을 위해 문제를 발견하고 문제를 해결하기 위한 데이터를 수집하는 과정이 우선적으로 수행된다. 이때 문제해결에 필요한 데이터를 수집하기 위한 방법으로 공공데이터 포털을 이용하고 있으며, 국내에서 생성되어 공유되고 출판되는 공공데이터는 모두 공공데이터 포털에서 간편하게 검색이 가능하다는 특징을 바탕으로 학교에서 수행되는 데이터사이언스 교육에서도 널리 활용되고 있는 편이다.

본 연구에서는 초등학교에서 학습자의 인지적 발달단계를 고려한 데이터사이언스 교육의 교수학습모형을 개발하는데 목표를 두고 있으며, 공공데이터를 활용하여 데이터를 수집하고 문제해결에 필요한 데이터를 찾기 위한 전략과 과정을 제시하고자 하였다. 특히, 공공데이터포털을 통해 적절한 데이터를 찾아 학교 교육에서 활용하는 관점으로 교수학습의 과정을 모델로 구성하여 일련의 절차로 제시하고자 하였다.

2. 이론적 배경

공공데이터포털은 공공기관에서 보유하고 있는 공공데이터를 국민들이 쉽게 접근하고 활용할 수 있도록 개방하고 있는 플랫폼으로서 공공데이터법이 시행됨에 따라 구축되어 제공되고 있다[15][16]. 데이터사이언스 교육을 위해 공공데이터포털에서 제공되고 있는 공공데이터를 활용하는 경우 검색이 기능을 활용하여 필요한 데이터를 발견하거나 혹은 카테고리 분류되어 제시하고 있는 데이터에서 필요한 데이터를 발굴할 수 있다. 신승기(2022)는 공공데이터포털에서 교육과 관련된 데이터를 키워드 검색 및 카테고리 분류를 통해 수집하여 토픽모델링을 통해 데이터의 특징을 살펴본 결과, 공공데

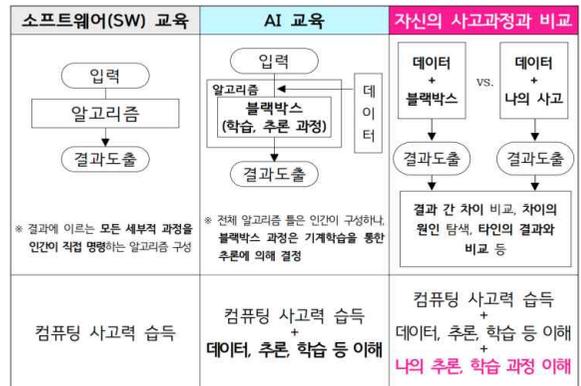
이더포털에서 제공하고 있는 데이터는 교육과 관련된 현황과 기관별로 제공되는 프로그램의 종류에 대한 내용으로 구성되어 있음을 살펴볼 수 있었으나, 교육과정 등과 같은 교육의 세부적인 운영 실태와 현황에 대해서는 데이터 활용이 제한적임을 설명하였다[23]. 공공기관에서 제공하는 데이터는 주로 현상에 대한 수치적 운영 현황에 대해 주로 제공되고 있다는 특징을 생각해볼 때 데이터사이언스 교육을 통해 일상생활에서의 문제를 해결하기 위한 데이터를 수집하는 과정에서는 별도로 데이터의 수집과 검색의 과정이 중점적으로 수행될 필요가 있음을 시사한다.

데이터사이언스 교육은 정보교육이 추구하는 컴퓨팅 사고력을 신장시키는 데 목표를 두고 있으며, 일상생활에서의 문제해결과정을 수행하여 창의적인 문제해결력 신장에 효과가 있다. 김용민(2020)은 초등학생을 대상으로 프로젝트기반 학습(PBL)을 수행하여 컴퓨팅사고력과 창의성의 변화에 긍정적인 결과를 분석하여 제시하였다 [7]. 구덕희와 김동진(2020)은 데이터사이언스 교육을 위한 교수학습 프로그램을 설계하였으며, 학생들의 문제해결력 신장을 위해 흥미 있는 주제를 선정해야 하며 데이터의 선정기준은 개인보다는 공공의 데이터를 선정하고 데이터의 양을 점차적으로 확대하는 문제해결 중심의 내용을 구성해야 문제해결력에 도움이 된다고 제시하였다[1]

학교 현장에서 AI교육의 도입을 통해 관심이 높아진 데이터사이언스 교육에 대한 흐름은 비단 우리나라의 현상뿐만이 아니다. 미국의 경우 세부 전공 및 과목으로 편성되어 운영된 데이터사이언스에 대하여 애리조나대학교를 비롯하여 8개 대학에서 관련 전공을 개설하여 전문가를 양성함을 살펴볼 수 있다[17]. 교육은 대학에서 전문가를 양성하는 과정뿐만 아니라 사회의 변화와 국가적 요구에 따라 모든 국민들에게 필요한 내용의 경우 초등학교에서부터 위계를 갖춘 체계를 통해 교육이 필요하다는 점에서 데이터사이언스 교육에 대한 교수학습방법 및 모형에 대한 연구가 필요하다.

정부에서 발표한 AI시대 교육정책 방향에서는 교육을 통한 인공지능 소양을 함양하기 위한 방법으로 학교 급별 위계를 갖춘 체계적인 정보교육을 실시해야 하며, SW교육과의 비교를 통해 AI교육에서는 데이터를 입력하여 AI알고리즘을 통해 예측과 분류를 수행하여 결과

를 도출하는 일련의 과정이라는 것을 (Fig. 1)과 같이 제시하고 있다[8]. 이는 인공지능 교육에서 데이터를 기반으로 문제를 해결하는 과정을 수행하는 것이 필요함을 의미하며, 데이터사이언스 교육으로서의 속성을 포함하고 있음을 나타낸다. 따라서 문제를 해결하는 과정에서 문제를 발견하고 문제해결에 필요한 데이터를 발굴하고 수집하여 정제하는 과정을 포함하는 일련의 교수학습과정이 필요함을 살펴볼 수 있다. 아울러, 데이터를 통해 얻게 되는 인사이트를 공유하고 타당도를 검증하기 위해 표와 그래프 등과 같은 형태로 시각화하고 의미를 도출하며 소통하는 과정을 포함하는 순차 및 비순차적 과정이 학교현장에 요구되고 있다.



(Fig. 1) Conceptual Framework of AI Education through Blackbox (Korean Government, 2020, p.31)

3. 연구 목적 및 연구 방법

국가수준 교육과정이 2022년도에 개정되면서 나타난 가장 큰 변화는 학생들에게 필요한 기초소양으로서 디지털 소양이 반영된 것이다. 디지털 소양은 디지털 대전환이라고하는 전 세계적인 변화에서 교육을 통해 핵심 역량으로 도출된 것으로서, 두 가지의 방법을 통해 기르도록 안내되어 있다. 첫 번째는 정보교육을 통해 디지털 소양을 함양하며 프로그래밍을 비롯한 피지컬컴퓨팅 등의 교수학습과정을 통해 컴퓨팅사고력을 기르도록 하는 것이다. 두 번째는 컴퓨팅사고력을 기반으로 모든 교과에서 디지털 소양을 활용하여 다양한 문제해결의 과정을 수행하도록 하여 융합적 역량을 기르도록 하는 것이다. 이는 정보교육을 통해 디지털 역량을 함양하고 다른

교과를 비롯하여 디지털 사회에서 다양한 문제상황에서 효과적으로 문제를 해결하도록 하는 역량을 기르는데 목표를 두고 있는 것이다. 이를 위해 정보과 교육과정에서는 2015 개정 교육과정과 비교하였을 때 가장 큰 변화로 인공지능교육이 내용 요소로서 반영되었으며, 컴퓨팅사고력을 기반으로 데이터를 활용하여 다양한 문제상황에서 문제해결능력의 과정을 수행하는 내용으로 제시되었다. 즉, 2022 개정 교육과정의 핵심은 디지털 역량을 강화하는 것이며, 인공지능교육을 반영하여 정보교육이 확대되는 부분을 고려한다면 데이터를 입력하여 분류하거나 미래를 예측하는 활동을 한다는 점에서 교육활동에서 데이터의 중요성은 매우 높다고 할 수 있다.

본 연구에서는 학교 현장에서 쉽게 접근 가능한 데이터 풀(Pool)로서 공공데이터를 활용하는 상황에서 교수학습과정에서 필요한 모델을 설계하는데 목표를 두고 있다. 특히 2022 개정 국가수준 교육과정 총론의 발표에 따라 인공지능 교육 및 소프트웨어 교육이 정보교과 및 모든 교과에서 활용되기 위한 방안이 제시되었다는 점에서 데이터 기반의 의사결정을 위한 교육을 위한 방향으로 모델을 구성하고자 하였다. 이는 인공지능 교육의 내용 요소중 하나로 데이터사이언스 교육을 통해 지도학습 및 비지도학습 기반의 데이터 활용 및 인사이트 도출의 방향에 대한 교수학습의 요구가 높아짐을 반영하기 위함이다. 아울러, 공공데이터를 활용한 데이터사이언스 교육을 위한 체계적인 교수학습모형이 필요하며 초등학교 현장에서 학습자의 인지적 발달단계를 고려한 과정 제시가 필요하다.

연구를 수행하는 과정에서 다음과 같은 절차로 모델을 설계하는 단계가 수행되었다. 첫째, 2022 개정 국가수준 교육과정 총론의 인공지능 교육과 정보교육에서 제시하고 있는 데이터사이언스 교육의 내용과 범위를 살펴보았다. 둘째, 데이터사이언스 교육을 위한 기저 사고과정으로서 컴퓨팅사고력(Computational Thinking)을 토대로 선행연구를 분석하며 교수학습단계 구성하였다. 특히, 데이터사이언스의 생명주기를 다룬 선행연구를 살펴보고 컴퓨팅사고력과 연계하여 모델을 구성하고자 하였다. 셋째, 공공데이터를 활용한 데이터 수집, 데이터 전처리, 학습, 모델링, 예측, 분류, 시각화 등의 과정을 선형 및 비선형 교수학습모형으로 구성하여 복합적인 교수학습의 과정이 활용될 수 있도록 설계하였다.

4. 개정 교육과정 총론의 디지털 교육 지원 체계

2022 개정 교육과정에서는 디지털 대전환이라는 사회적 변화와 시대적 요구를 반영하기 위해 교육과정 개정의 중점으로 “인공지능 기술 발전에 따른 디지털 전환”을 첫 번째 배경으로 제시하고 있다[12]. 이는 미래를 살아갈 학생들이 해결해야 할 문제들은 불확실성이 높은 것들이 대다수이며, 고정된 지식으로 해결할 수 있는 것이 아니라 다양한 상황에서 문제를 해결할 수 있는 능동적인 태도와 주도성을 함양하기 위함을 밝히고 있다[12]. 따라서 교육과정의 총론의 5쪽에서는 다음과 같이 교육과정 구성의 중점사항을 제시하고 있다[12].

모든 학생이 학습의 기초인 언어·수리·디지털 기초소양을 갖추도록 하여 학교 교육과 평생 학습에서 학습을 지속할 수 있게 한다.

학습의 기초소양을 기존의 3Rs를 근간으로 하는 읽기, 쓰기, 셈하기로 대표되는 언어와 수리 역량에서 확대하여 디지털 역량을 기초소양으로 정의하고 있음은 2022 개정 교육과정의 대표적인 변화 중 하나라고 할 수 있다. 이를 위해 교육과정 총론에서는 언어·수리·디지털 기초소양을 모든 교과에서 함양할 수 있도록 수업을 설계하도록 제시하고 있다[12].

학습환경의 측면에서는 디지털 기반 학습이 가능한 교육 공간과 환경을 조성해야 하며, 다양한 지능정보기술 및 도구를 활용하여 효율적인 학습을 지원해야함을 제시하고 있다. 이를 위해 국가 수준에서는 디지털 교육 환경 변화에 부합하는 교원의 역량 함양을 지원해야 하며, 교육청 수준에서는 수업 설계·운영과 평가에서 다양한 디지털 플랫폼과 기술 및 도구를 활용하기 위한 시설과 설비 및 기자재가 필요함을 기술하고 있다[12].

데이터사이언스 교육은 문제를 발견하고 문제해결의 아이디어를 구성하고 검증하기 위해 관련된 데이터를 수집하고 분석하는 과정에서 컴퓨팅 장비를 활용한 자동화 시스템을 이용하고, 수학적 이론과 원리를 기반으로 데이터의 적절성과 인사이트를 발굴하며 이를 말과 글로 표현하고 공유하게 된다는 점에서 2022 개정 교육과정의 기초소양인 언어·수리·디지털 기초소양은 데이터사이언스 교육의 핵심역량과 동일하다. 다만, 데이터

사이언스 교육에서는 이와 같은 기초소양을 기반으로 도메인 지식과 역량을 활용하여 다양한 문제상황에서 적용하고 문제해결의 아이디어를 구성하는 것이 중요한 관점이라는 점에서 보다 확장된 개념으로 적용하는 범주와 의미를 갖는다. 2022 개정 교육과정에서 제시된 디지털 소양을 함양하기 위한 교육과정 개정의 중점과 배경 및 지원 체계는 인공지능 시대의 불확실한 문제를 해결하는 아이디어를 구성하기 위한 능동적이고 주도적인 사고과정을 수행하도록 한다는 점에서 데이터사이언스 교육의 체계를 지원하는 핵심적인 내용과 변화라고 할 수 있다. 이를 정리하면 <Table 1>과 같이 살펴볼 수 있다.

<Table 1> Digital Education in 2022 National Curriculum[12]

구분	디지털교육요소
교육과정 개정 주요 배경	인공지능 기술 발전에 따른 디지털 전환, 감염병 대유행 및 기후·생태환경 변화, 인구 구조 변화 등에 의해 사회의 불확실성이 증가하고 있다.
교육과정 구성의 중점	디지털 전환, 기후·생태환경 변화 등에 따른 미래 사회의 불확실성에 능동적으로 대응할 수 있는 능력과 자신의 삶과 학습을 스스로 이끌어가는 주도성을 함양한다. 모든 학생이 학습의 기초인 언어·수리·디지털 기초소양을 갖추 수 있도록 하여 학교 교육과 평생 학습에서 학습을 지속할 수 있게 한다.
교수·학습	교과의 깊이 있는 학습에 기반이 되는 언어·수리·디지털 기초소양을 모든 교과를 통해 함양할 수 있도록 수업을 설계한다. 다양한 지능정보기술 및 도구를 활용하여 효율적인 학습을 지원할 수 있도록 디지털 학습 환경을 구축한다.
교육 환경 조성	국가 수준 디지털 교육 환경 변화에 부합하는 미래형 교수·학습 방법과 평가체제 구축을 위해 교원의 에듀테크 활용 역량 함양을 지원한다. 수업 설계·운영과 평가에서 다양한 디지털 플랫폼과 기술 및 도구를 효율적으로 활용할 수 있도록 시설·설비와 기자재 확충을 지원한다.

5. 초등학교 정보교육의 데이터사이언스 교육

2022 개정 국가수준 교육과정에서 정보교육은 이전 교육과정인 2015 개정 국가수준 교육과정과 비교하여 디지털 인재 양성을 위해 시수가 2배로 확대되었다. 2015 개정 국가수준 교육과정에서 초등학교는 17시간, 중학교는 34시간으로 편성되어 있었던 것을 2022 개정 국가수준 교육과정에서는 초등학교 34시간 이상, 중학교 68시간 이상 운영하도록 제시하고 있다[12].

초등학교 정보교육은 2015 개정 국가수준 교육과정에서 실과교과에서 17시간이 편성되어 있었으며, 2022 개정 국가수준 교육과정에서는 실과교과의 17시간과 함께 새롭게 편성되는 학교자유시간 등을 활용하여 17시간을 확보하여 총 34시간 이상 이수하기 위해 학교교육과정을 편성하도록 제시하고 있다. 따라서 2022 국가수준 교육과정에서 초등학교 정보교육은 실과교과에서 하나의 영역으로 17시간이 편성되어 교육과정이 구성되어 있으며, 부록으로 추가 17시간의 성취기준과 성취기준 해설이 작성되어 있다. 즉, 실과교과의 내용체계에는 실과에서 이수되는 17시간의 내용만 편성되어 있으며, 추가 17시간의 내용은 시수의 하한기준과 성취기준은 제시되어 있으나, 내용체계에서 반영되는 영역과 중학교와의 연계가 미흡하게 제시되어 있다고 할 수 있다.

<Table 2>는 2022 개정 국가수준 교육과정에 편성된 실과교과에 제시된 초등학교 정보교육의 성취기준이며, 데이터사이언스 교육과 직접적으로 연계된 성취기준을 표기한 것이다. 초등학교 실과교과에 편성된 정보교육의 성취기준은 “(5) 디지털 사회와 인공지능” 영역의 5개로 편성되어 있으며, 데이터의 이해 및 인공지능에서 활용하기 위한 관련된 내용을 살펴볼 수 있다.

<Table 2> Data Science Education in Elementary School[13]

성취기준	데이터사이언스교육
[6월05-04] 디지털 데이터와 아날로그 데이터의 특징을 이해하고, 인공지능에 활용할 수 있는 데이터의 유형이나 형태를 탐색한다.	▶ 데이터의 의미 이해 ▶ 데이터의 유형과 형태 탐색
[6월05-05] 인공지능이 만들어지는 과정을 체험하고, 인공지능이 사회에 미치는 영향을 탐색한다.	▶ 기계학습이 적용된 인공지능 도구 체험 ▶ 기계학습의 원리 이해

<Table 3>은 2022 개정 국가수준 교육과정에 편성된 실과교과의 부록으로 제시되어 학교자율시간 등에서 활용하는 경우 참고하는 초등학교 정보교육의 성취기준이며, 데이터사이언스 교육과 직접적으로 연계된 성취기준을 표기한 것이다. 초등학교 실과교과의 부록으로 편성된 정보교육의 성취기준은 모두 7개로 구성되어 있으며, 실과교과에 편성된 성취기준과 연결되어 제시되어 있다. <Table 3>에 제시된 두 개의 성취기준은 실과교과의 [6설05-04]와 연결되는 내용으로, 데이터의 의미를 이해하고 유형과 형태를 탐색하며 생활 속에서의 다양한 데이터를 이해하고 목적과 의미를 살펴보는 활동으로 연결된다. 생활 속에서 수집된 데이터를 활용하여 인공지능으로 분류를 수행하여 데이터의 공통되는 부분을 찾는 알고리즘을 체험하고 내면화하며, 이를 토대로 기계학습의 개념을 활용하여 인공지능 도구를 체험하여 그 원리를 이해하도록 구성되어 있다.

<Table 3> Data Science Education in Elementary School[13]

성취기준	데이터사이언스교육
[06자율-5] 생활 속에서 접하는 다양한 데이터가 서로 다른 의미를 갖고 있음을 파악하는 활동을 경험한다.	▶ 생활 속 데이터의 이해 ▶ 데이터의 목적과 의미
[06자율-6] 인공지능이 데이터의 공통되는 부분을 찾아 분류하는 과정을 체험한다.	▶ 인공지능의 분류과정 ▶ 내면화

6. 컴퓨팅사고력 기반 데이터사이언스 생명주기 분석

데이터사이언스 교육은 인공지능교육의 범주에서 데이터를 활용하여 문제를 해결하고 문제해결의 아이디어를 얻기 위해 수행되는 일련의 교수학습과정이다. 이를 위해 데이터를 활용하여 예측과 분류를 수행하고 인사이트를 얻어 문제해결의 인사이트를 발굴하게 되는 선형 혹은 비선형의 과정이라고 할 수 있다. 컴퓨팅사고력(Computational Thinking)은 Papert(1980)에 의해 처음 용어가 사용되었으며 현재의 개념은 Wing(2006)에 의해 정립되었다. Wing(2006)은 추상화(Abstraction)와 자동화(Automation)가 핵심과정이라고 설명하고 있으며, 문제해결의 아이디어를 알고리즘(Algorithm)으로 구성

한다고 제시하였다. Wing(2019)은 데이터사이언스 교육을 <Table 4>와 같이 8개의 단계로 제시하고 있다.

<Table 4> Wing(2019)'s Data Life Cycle[26]

Stages	Process
Generation	Data generating by people and sensors.
Collection	Deciding what to collect defines a filter on the data we generate.
Processing	Everything from data cleaning, data wrangling, and data formatting to data compression
Storage	Data storage in efficient storage, and data encryption, for secure storage
Management	Create and use different kinds of metadata to maximize our ability to access and modify the data for subsequent analysis. for those dimensions of heterogeneity. which are: ① structured and unstructured data, ② data of many types (e.g., text, audio, image, video), ③ data that arrive at different velocities
Analysis	All the computational and statistical techniques for analyzing data is being conducted to gain knowledge or insights, build classifiers and predictors, or infer causality
Visualization	Present results in a clear and simple way along with human visual perception to convey the results of data analysis
Interpretation	Tell a story explaining the picture's context, point, implications, and possible ramifications

데이터의 생성(Generation)은 사람들의 디지털 생활로부터 생성되기도 하며 사물인터넷(IoT) 장비들의 센서들을 통해서 데이터가 만들어지기도 한다. 생성된 데이터들은 수집(Collection)의 과정을 거치면서 모든 데이터가 수집되는 것이 아니라 어떤 데이터가 수집되게끔 할 것인지 거름망(filter)를 선별하여 데이터가 생성되는 과정에서 자동으로 수집되며 저장(Storage)의 단계로 이어진다. 저장된 데이터는 관리(Management)의 단계를 통해 정형데이터 및 비정형 데이터 등을 구별하여 데이터 사용의 가능성을 높이기 위한 메타데이터를 생

성하게 된다. 이어서 분석(Analysis)의 과정은 데이터의 생명주기에서 가장 중요한 과정이라고 Wing(2019)은 설명하고 있으며, 인공지능, 데이터마이닝, 기계학습 등을 활용하여 예측과 분류를 통해 지식과 인사이트를 발굴하는 단계라고 설명하고 있다. 분석된 데이터는 시각화(Visualization)과정을 통해 그래프와 그림으로 표현하여 데이터의 분석 결과를 시각적으로 인지하여 빠르게 정보를 전달하도록 하며, 해석(Interpretation)의 과정을 통해 시각화된 자료를 설명하는 과정이 수행된다. 데이터의 분석과정은 이와 같은 8개의 과정으로 수행되지만, 사람들은 데이터 분석 결과를 바탕으로 정책을 결정하거나 미래를 예측하여 효과적인 의사결정을 하는 중요한 자원으로 활용하게 된다. Wing(2019)은 데이터 생명주기를 설명하면서 각각의 단계는 선형적 연결성을 갖는 것이 아니라 필요한 단계를 집중적으로 수행하거나 순서를 달리할 수 있음을 함께 설명하고 있다.

Microsoft(2022)는 데이터사이언스 생명주기를 제시하기 위하여 TDSP(The Team Data Science Process Lifecycle)라는 이름의 데이터사이언스에 대한 프레임워크를 제시하고 있으며 데이터마이닝을 비롯하여 데이터사이언스 기반의 프로젝트 수행에 필요한 과정을 <Table 5>와 같이 5개의 단계로 제시하고 있다.

<Table 5> Microsoft(2022)'s Data Life Cycle[9]

Stages	Process
Business Understanding	▶ Define objectives ▶ Identify data sources
Data Acquisition and Understanding	▶ Ingest the data into the target analytic environment. ▶ Explore the data to determine if the data quality is adequate to answer the question. ▶ Set up a data pipeline to score new or regularly refreshed data.
Modeling	▶ Feature engineering ▶ Model training ▶ Determine if the model is suitable for production
Deployment	▶ Deploy the model and pipeline to a production
Customer Acceptance	▶ System validation ▶ Project hand-off

문제발견(Business Understanding)의 과정에서는 데이터를 기반으로 해결해야 하는 목표와 문제를 정의하고 필요한 데이터를 발굴하는 과정으로 제시되어 있다. 이어서 수행되는 데이터 수집 및 탐색(Data Acquisition and Understanding)의 단계에서는 데이터를 수집하고 수집된 데이터의 질적 평가를 통해 문제해결에 도움이 되는지 판단하며 주기적인 데이터 수집과 평가가 수행될 수 있도록 데이터 파이프라인을 설정하는 과정이 수행된다. 수집된 데이터를 바탕으로 모델링(Modeling)의 과정에서는 필요한 인공지능 기술을 파악하고 모델을 학습시키며 생성된 모델이 문제해결에 적합한지 확인하는 과정이 수행된다. 학습된 모델은 적용(Deployment)의 과정을 통해 데이터 기반의 문제해결과정에 실제 적용되며 수락(Customer Acceptance)의 과정을 통해 일련의 데이터사이언스 기반의 문제해결과정이 종료된다.

ICPSR(Inter-University for Political and Social Research)는 1962년에 설립되었으며 미시간대학교 내에 위치하고 있는 컨소시엄이다[4]. 전 세계의 750개 이상 연구기관과 학술단체가 가입되어 있으며 25만개 이상의 사회과학에 대한 데이터를 유지 및 관리하고 있다. <Table 6>은 ICPSR (2020)에서 발간한 데이터 생명주기를 나타낸 것으로서 6개의 단계로 구성된다.

<Table 6> ICPSR (2020)'s Data Life Cycle[3]

Stages	Process
Proposal Development and Data Management	▶ Data Description ▶ Review of Existing Datasets ▶ Format and Metadata
Project Start-up	▶ Make decisions about documentation form and content ▶ Conduct pretests and pilot tests of materials and methods
Data Collection and File Creation	▶ Dataset Creation and Integrity for Quantitative Data ▶ Confidentiality in Qualitative Data
Data Analysis	▶ Manage datasets and work files ▶ Set up appropriate file structures ▶ Back up data and documentation
Preparing Data for Sharing	▶ Address disclosure risk limitation ▶ Determine file formats to deposit ▶ Complete relevant forms
Depositing Data	▶ Comply with dissemination standards and formats

문제발견 및 데이터 수집 계획 수립의 단계라고 할 수 있는 Proposal Development and Data Management에서는 기존에 존재하는 데이터셋을 살펴보고 데이터의 종류를 파악하여 데이터의 형식과 메타데이터에 대해서 내용을 살펴보는 과정이 수행된다. 해결이 필요한 문제에 대하여 계획을 수립하는 단계로서 Project Start-up이 수행되며, 이미 갖고 있거나 수집되어 있는 데이터를 활용하여 파일럿 테스트를 통해 데이터의 종류에 따른 문제해결의 가능성을 살펴본다. 이어서 데이터 수집 및 파일 생성(Data Collection and File Creation)의 과정을 통해 문제해결에 필요한 데이터를 수집하고 수집된 데이터의 일관성을 유지하기 위해 변수와 라벨 및 그룹의 관리와 함께 누락된 데이터 관리 등을 수행하게 된다. 특히 양적 데이터와 질적 데이터는 특징에 따라 구분하여 데이터 수집 및 파일 생성이 이루어진다. 수집된 데이터를 바탕으로 실제 데이터 분석(Data Analysis)이 수행되며 데이터셋의 관리와 함께 파일 구조 및 백업 데이터 관리 등이 병행되어 이루어진다. 분석된 결과를 공유하기 위하여(Preparing Data for Sharing) 파일의 형태를 정의하며 공개의 범위를 결정하게 된다. 공유된 데이터는 최종적으로 출판과 공유를 위한 형태가 결정되며(Depositing Data), 데이터를 보관하거나 이차데이터로 생성하기 위한 여부를 결정하게 된다.

Stodden(2020)은 ACM(Association for Computing Machinery)의 기고문을 통해 데이터사이언스의 생명주기를 활용하여 문제를 해결하는 과정에 <Table 7> 같은 과정이 필요함을 제시하였다[25]. 데이터를 수집하고 문제해결에 필요한 데이터를 추출하며 예측 혹은 분류를 위한 인공지능 모델 중 설명력이 높은 것을 선택하여 타당도를 검증하며 문제를 해결하는 과정으로 제시하고 있다. 이는 Wing(2008)의 컴퓨팅사고력에 대한 정의에서 추상화를 통한 데이터 추출 및 인공지능 모델로 학습시키고 문제를 해결하는 자동화 과정이 중심이 되는 일련의 데이터사이언스의 절차이다.

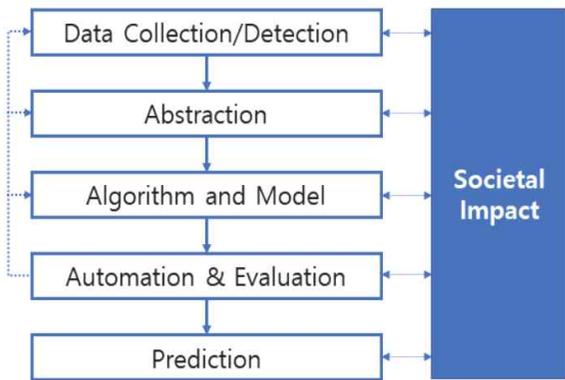
<Table 7> Stodden(2020)'s Data Science Life Cycle[25]

Obtain Data → Data Preparation → Feature Selection/Model Estimation → Cross-tests and Validation → Publication and Archiving

Kadijevich(2018)은 데이터사이언스 교육을 통해 컴퓨팅사고력을 기르는 과정이 필요함을 제시하였으며, 데이터 시각화를 통해 상호작용이 가능한 그래프를 구성하여 데이터가 변화됨에 따라 변화되는 모습을 살펴 보면서 문제해결과정을 수행하는 것이 필요하다고 제시하고 있다. 특히 모델링을 통한 데이터 분석을 통해 실제적인 문제해결과정을 체험하도록 해야하며, 컴퓨팅사고력을 기르기 위해 컴퓨팅이 기반이 되는 환경(Computational Environment)에서 데이터를 기반으로 문제를 해결하는 내용이 수행되어야 한다고 제시하였다. 즉, 상호작용이 가능한 그래프를 기반으로 수치적 데이터를 활용하여 데이터의 값이 변화되는 과정에서 이상적인 방향과 전략을 도출하는 것이 컴퓨팅사고력 신장에 효과적인 이른바 Computational Strategies라고 제시하였다[5].

신승기(2019)는 인공지능교육을 위한 프레임워크를 제시하기 위하여 컴퓨팅사고력을 기반으로 인지적학습 환경을 제공하기 위한 관점으로 교수학습모형을 설계하여 제시하였다. 인지적 학습환경 프레임워크는 컴퓨팅 사고력 모델을 기반으로 인공지능 교육을 위한 사고과정 측면에서의 절차를 분석하여 Agency(인지적 학습보조), Abstracting(추상화), Modeling(모델링)의 단계로 구성되어 있으며, 5개 단계의 교수학습모형이 설계되어 제시되었다[20]. 인지적 학습환경 프레임워크에 대하여 서성재와 김철(2021)은 예비 교원을 대상으로 인공지능 교육을 실시하기 위해 신승기(2019)의 교수학습모형을 활용하여 컴퓨팅사고력 신장에서 효과성을 검증하였다[19]. 김빛나라와 신승기(2023)는 데이터를 활용한 인공지능 교육에서 인지적 학습환경 프레임워크를 적용하여 수업에 적용하여 질적연구를 통해 학생들의 데이터 기반의 문제해결력이 길러짐을 살펴볼 수 있었다. 이는 신승기(2019)가 제시한 인지적 학습환경 프레임워크 및 인공지능교육 교수학습모형의 타당성을 뒷받침한다[6].

<Fig. 2>는 신승기(2019)에 제시된 인공지능교육을 위한 교수학습 모델로서 컴퓨팅사고력을 근간으로 설계된 일련의 과정이다. 추상화와 자동화 과정을 중심으로 문제해결의 알고리즘을 통해 수집된 데이터를 정제하고 알고리즘을 통해 모델링을 수행하며 결과를 도출하고 예측하는 과정을 수행하게 되며, 사회적 영향을 모든 단계에서 고려하여 진행되는 선형 및 비선형의 과정이다.



<Fig. 2> Shin(2019)'s AI Learning Instructional Model[20]

컴퓨팅사고력을 기반으로 데이터사이언스 교육을 실시하기 위한 교수학습모형의 프레임워크는 <Table 8>과 같이 살펴볼 수 있다. Computational Thinking Model(컴퓨팅사고력 모델)은 미국의 CSTA(Computer Science Teachers Association)에서 개발된 단계로서[2],

2015 개정 교육과정을 통해 소프트웨어 교육이 도입되면서 문제해결과정의 기저 단계로 제시된 컴퓨팅사고력의 개념과 단계에 대한 맥락이 일치한다[10][11][18][20][24]. 이를 토대로 신승기(2019)는 인공지능교육을 위한 인지적 학습환경을 설계하기 위한 프레임워크를 설계하였으며, <Fig. 2>에 제시된 내용과 일치한다. 또한 <Table 8>에 제시된 인공지능교육을 위한 교수학습 단계에서 제시된 단계가 Agency(인지적 학습보조), Abstracting(추상화), Modeling(모델링)의 과정에 대한 세부 과정과 내용으로 구성된 것이다[20][21].

본 연구에서 설계하고자 하는 데이터사이언스 교육을 위한 교수학습모형을 설계하기 위하여 공공데이터를 활용하고 컴퓨팅사고력을 신장시키기 위한 관점으로 설계하기 위하여 데이터사이언스의 생명주기를 단계별로 살펴해보면서 데이터의 생성과 수집에서 분석과 시각화 및 일반화에 이르는 과정에 대해 대표적인 연구 결과를 살펴보고 기존의 컴퓨팅사고력 모델과 인공지능교육 교수학습모형과의 비교를 통해 내용을 살펴보았다.

<Table 8> The Framework for Data Science Education through Computational Thinking

Computational Thinking Model	AI Instructional Model of Shin(2019)	Data Science Life Cycle		
		Wing(2019)	ICPSR(2020)	Microsoft(2022)
		Data Generation	Review of Existing Datasets	Identify Data Sources
			Pilot Tests	
Data Collection	Data Collection	Data Collection	Data Collection and File Creation	Data Acquisition
Data Analysis	Data Detection	Data Processing	Data Analysis	Data Understanding
Data Representation	<ul style="list-style-type: none"> Reasoning Recursion Repetition 	Data Storage		
Problem Decomposition				
Abstraction	Abstraction	Data Management		Data Pipeline
Algorithms and Procedures	Algorithms & Modelling		Preparing Data for Sharing	Model Training
Automation	Automation & Evaluation	Data Analysis		
Simulation	<ul style="list-style-type: none"> Modulating Training & Evaluation 	Visualization		Deployment
Parallelization		Prediction	Interpretation	Depositing Data

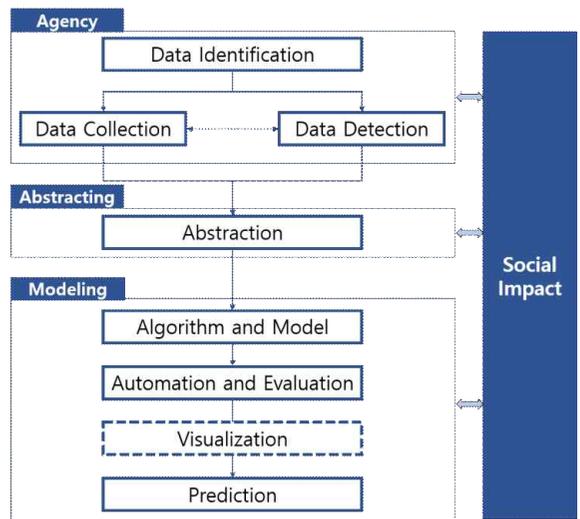
기존의 교수학습모형은 문제의 발견으로부터 데이터의 수집이 첫 번째 단계로 수행되고 있는 반면, 데이터 사이언스 생명주기에 따르면 데이터의 생성의 출처를 살펴보고 기존에 갖고 있는 데이터셋을 분석하며 필요한 데이터를 수집하기 위한 방법을 찾는 과정에 대해 선행적 단계가 포함되고 있음을 알 수 있다. 특히, ICPSR(2020)에서는 한 단계 나아가서 현재 갖고 있는 데이터를 통해 문제를 해결할 수 있는 아이디어와 범주를 살펴보기 위한 파일럿 테스트를 수행하고 데이터를 수집한다는 내용도 살펴볼 수 있다. 아울러, 데이터 분석의 결과를 다른사람들과 공유하기 위한 시각화 및 표현의 과정이 포함되어 있음도 특징적인 과정이라고 할 수 있으며, 결과에 대한 해석과 설명이 단계에 제시되어 있음을 알 수 있다.

7. 공공데이터활용 데이터사이언스 교육 모형 설계

2022 개정 교육과정의 도입에 따라 디지털 소양이 기초 필수역량으로 총론에서 제시되면서 정보교육을 통한 역량강화를 비롯하여 모든 교과에서 디지털 역량을 기르기 위한 내용이 반영되었다. 정보교육에 대한 내용요소 및 성취기준의 가장 큰 변화 중 하나는 인공지능 시대에서의 역량을 강화하기 위한 내용으로 데이터사이언스를 기반으로 데이터 수집 및 분석을 기반으로 결과를 해석하고 실생활에서의 문제를 해결하는 내용이 컴퓨팅 사고력을 기반으로 나타나있다.

정보교육에서는 디지털 소양을 비롯하여 이를 자동화하는 SW교육과 함께 AI교육에서 제시하는 인공지능의 알고리즘을 활용하여 예측과 분류를 수행하는 핵심역량을 기르는 내용이 교육과정에 제시되어 있다. 이와 더불어 모든 교과에서도 디지털 역량을 기반으로 불확실한 미래의 문제해결역량을 기르기 위한 활동이 반영될 수 있도록 안내하고 있다.

일상에서의 다양한 문제를 해결하기 위해 문제를 발견하고 이를 해결하기 위한 데이터를 수집하며 문제해결의 과정을 수행하면서 데이터 분석결과를 시각화하는 활동이 요구되고 있다. 특히, 데이터를 수집하기 위하여 공공데이터법을 기반으로 운영되고 있는 공공데이터포털을 비롯하여 다양한 공공데이터를 활용하여 일상에서의 문제를 해결하기 위한 노력이 추진되고 있다.



<Fig. 3> Instructional Model for Data Science Education

<Fig 3.>에 제시된 데이터사이언스 교육을 위한 교수학습모형은 인공지능 교육을 위한 인지적 학습환경의 프레임워크인 Agency(인지적학습보조), Abstraction(추상화), Modeling(알고리즘구현)을 기반으로 <Table 8>에 제시된 데이터사이언스 생명주기에서 도출된 컴퓨팅 사고력과 연계된 문제해결의 과정을 토대로 교수학습모형이 구성되었다. 모든 과정에서 사회적 영향을 고려하며, 데이터 식별을 통한 수집과 발견이 강화되었다.

<Table 9>에 나타난 내용과 같이 공공데이터를 활용하는 데이터사이언스 교육의 교수학습모형의 단계가 제시되었다. 데이터 식별(Data Identification)의 과정에서는 해결이 필요한 문제를 발견하는 것이 선행 되어야하며, 문제해결에 필요한 데이터의 출처를 살펴보고 기존의 데이터셋 중에 활용가능한 것이 있는지 살펴보는 과정이 제시되어 있다. 데이터 수집(Data Collection)의 과정과 데이터 발견(Data Detection)의 과정은 비선형의 과정으로 구성되며, 문제를 해결하기 위해 요구되는 데이터를 수집하고 이를 파일로 정리하는 과정과 함께 수집된 데이터 혹은 기존의 데이터가 문제해결에 필요한 적절한 수준의 양과 데이터인지를 판단하는 과정이 수행된다. 추상화(Abstraction)의 과정에서는 문제해결을 위해 수집된 데이터에 대하여 반드시 해결이 필요한 데이터만 간결하게 정리하는 과정이 수행된다. 알고리즘과 모델링(Algorithm and Model)의 과정에서는 데이터

를 활용하여 문제해결 모델을 생성하기 위해 적절한 알고리즘을 선택하는 활동이 수행된다. 자동화 및 평가(Automation and Evaluation)의 단계에서는 문제해결에 필요한 데이터를 분석하기 위해 데이터 분석을 수행하고 모델의 설명력을 평가하는 과정이 수행된다. 시각화(Visualization)의 단계에서는 데이터 분석의 결과를 표와 그래프와 같은 시각화 과정을 수행하며 데이터 분석의 파이프라인의 설계 결과를 공유하도록 한다. 예측(Prediction)의 과정에서는 데이터분석의 결과를 해석하고 문제해결을 예측하기 위해 사용되는 문제해결의 모형으로 적용하는 활동이 수행된다. 사회적 영향(Social Impact)은 모든 단계에서 수행되며 데이터의 식별과 수집 및 분석과 시각화와 예측과 설명의 모든 과정에서 사회적인 영향을 고려하여 문제해결과정에서 활용될 수 있도록 한다.

공공데이터를 활용한 데이터사이언스 교육 교수학습 모형은 실생활에서의 문제를 해결하기 위해 필요한 데이터를 발견하는 과정으로부터 시작되어 데이터를 수집하고 이를 추상화 과정을 통해 간결하게 나타내며 인공지능 모델을 통해 문제해결의 모델을 설계하고 적용하여 결과를 공유하기 위한 활동이 수행되는 내용으로 구성되어 있다. 특히, 데이터를 수집하는 과정에서 기존의 데이터를 리뷰하고 요구되는 데이터를 식별하며 수집된 데이터의 속성에 대한 이해가 선행되어야 한다.

8. 결론 및 제언

교육과정의 개정을 비롯하여 디지털 대전환 시대에서 디지털 역량을 강화하기 위한 교육의 변화는 사회적·시대적으로 요구되고 있는 숙명이다. 인공지능 사회의 도래에 따라 예측이 더욱 불가능한 미래 사회에서 국가적 인재로 성장하고 나아가 세계를 선도하는 역량을 기르기 위해 불확실한 문제상황을 해결할 수 있는 방법을 배우는 과정이 필요하다. 이는 2022 개정교육과정 총론에서 개정의 중점으로 제시할 만큼 과거의 경험과 고전적 지식이 미래의 문제해결과정에서는 다양성과 창의적 문제해결이라는 역량을 통해 발휘되어야 하는 여건이다.

2022 개정 교육과정에서의 변화는 정보교육의 확대를 통해 디지털 소양을 기르고 컴퓨팅사고력기반의 문제해결력을 문제해결의 과정에서 향상시킬 수 있도록 제시하고 있다. 특히 인공지능시대의 도래에 따라 인공지능의 알고리즘을 이해하고 데이터를 학습하여 예측과 분류를 통해 문제해결력을 기르는 내용이 반영되어 있다. 아울러, 모든 교과에서 디지털 역량을 활용하여 다양한 도메인영역에서의 문제해결과정을 수행할 수 있도록 방향이 제시되어 있다.

데이터를 기반으로 문제를 해결하는 과정을 활동을 통해 배우는 데이터사이언스 교육에서는 데이터수집의 어려움으로 공공데이터를 활용하여 문제해결과정에 적

<Table 9> Description of Instructional Model for Data Science Education

Domain	Instructional Model	Description
Agency	Data Identification	Review the existing datasets and identify the data sources for problem-solving
	Data Collection	Collect the data and file the dataset to supply the materials to solve the problem
	Data Detection	Understand the characteristics of dataset and explore the data quality if data is adequate to answer the problem
Abstracting	Abstraction	Abstracting the data to extract the core data from dataset to solve the problem
Modeling	Algorithm and Model	Choose the appropriate algorithm for problem-solving modeling using the data
	Automation and Evaluation	Training the data analysis model and analysis the data to solve the problem
	Visualization	Visualize of the results of data analysis to share the deploy the pipeline
	Prediction	Apply the model for problem-solving to predict and interpret the result of data analysis
	Social Impact	Consider the social impact in all processes of data analysis.

용하는 것이 일반적이다. 이미 우리나라를 비롯한 전세계적으로 공공데이터법을 통해 공공기관에서 수집되고 보관되는 데이터는 국민들이 쉽게 활용할 수 있도록 온라인 포털을 통해 데이터를 제공하고 있다. 그러나 공공 데이터를 활용하는 과정에서 단지 데이터를 검색하여 필요한 데이터를 발굴하는데 초점을 두는 것이 아니라, 문제상황에 맞는 데이터를 수집하고 속성과 특징을 분석하여 필요한 데이터를 식별하는 것이 선행되어야 한다. 특히 데이터 생명주기에서 제시된 일련의 과정인 데이터를 수집하고 분석하며 결과를 공유하는 과정에 대해 교육이라는 환경에서 컴퓨팅사고력을 기반으로 문제 해결력을 기르는 모형으로서의 단계가 정립되어 교육현장에 안내될 필요가 있다. 2022 개정 교육과정이 확정고시됨에 따라 디지털 인재 양성을 위한 데이터 기반 교육(Data-driven Education)의 과정에서 본 연구의 교수 학습모형이 효과적으로 활용될 수 있을 것이다.

참고문헌

- [1] Koo, D., & Kim, D. (2020), Data science education program based on problem solving learning, *The Journal of Korea Elementary Education*, 31, 203-215.
- [2] Computer Science Teachers Association (CSTA) and the International Society for Technology in Education (ISTE) (2011), *Computational Thinking Teacher Resources*. Second Edition, Retrieved from https://c.ymcdn.com/sites/www.csteachers.org/resource/resmgr/472.11CTTeacherResources_2ed.pdf
- [3] ICPSR (2020), *Guide to Social Science Data Preparation and Archiving*. 6th edition, Retrieved from <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>
- [4] ICPSR (2023), *About ICPSR*, Retrieved from <https://www.icpsr.umich.edu/web/pages/about/>
- [5] Kadijevich, D. M. (2018), *Cultivating computational thinking through data practice*. In *Open Conference on Computers in Education*, Springer, 24-33.
- [6] Kim, B., & Shin, S. (2023), *Designing a data science and curriculum convergence class to develop computational thinking*, *The Korean Association of Information Education Research Journal*, 1(1), 177-182.
- [7] Kim, Y. (2020), *The effects of PBL-based data science education classes using app inventor on elementary student computational thinking and creativity improvement*. *Journal of The Korean Association of Information Education*, 24(6), pp. 512-562.
- [8] Korean Government (2020), *AI era, direction of educational policy and core tasks*. Retrieved from <https://www.korea.kr/archive/expDocView.do?docId=39237>
- [9] Microsoft (2022), *The team data science process lifecycle*. Retrieved from <https://learn.microsoft.com/en-us/azure/architecture/data-science-process/lifecycle>
- [10] Ministry of Education (2015), *Software education instructional guidance*.
- [11] Ministry of Education (2015), *Informatics curriculum. #2015-74 (Annex 10)*.
- [12] Ministry of Education (2022), *2022 revised curriculum. #2022-33 (Annex 1)*.
- [13] Ministry of Education (2022), *2022 revised informatics curriculum. #2022-33 (Annex 10)*.
- [14] Ministry of Education (2022), *Comprehensive digital talent development plan*. Press Release. Retrieved from <https://www.moe.go.kr/boardCnts/viewRenew.do?boardID=72770&boardSeq=92584&lev=0&searchType=null&statusYN=W&page=4&s=moe&m=0315&opType=N>
- [15] Ministry of Safety and Public Administration (2013), *Create jobs by loosening the shackles of public data!*. Retrieved from https://mois.go.kr/frt/bbs/type010/commonSelectBoardArticle.do?bbsId=BBSMSTR_00000000008&nttId=30053
- [16] Ministry of Safety and Public Administration (2013), *Major work plan of the Ministry of Safety and Public Administration in 2013*. Retrieved from <https://www.korea.kr/archive/expDocView.do?docId=33763>

[17] Park, H. (2022), Degree programs in data science at the school of information in the states. *Journal of Korean Library and Information Science Society*, 53(2), 305-332.

[18] Park, P., & Shin, S. (2019), A Study on the Instructional System and Curriculum Design to Evolve the Software Education in Elementary School, *Journal of The Korean Association of Information Education*, 23(3), 273-282.

[19] Seo, S., & Kim, C. (2021), Analysis of understanding of prospective teachers' computational thinking on artificial intelligence education. *Journal of The Korean Association of Information Education*. 25(1), 123-134.

[20] Shin, S. (2019), Designing the instructional framework and cognitive learning environment for artificial intelligence education through computational thinking. *Journal of The Korean Association of Information Education*, 23(6), 639-653.

[21] Shin, S. (2020), Designing the framework of evaluation on learner's cognitive skill for artificial intelligence education through computational thinking, *Journal of The Korean Association of Information Education*, 24(1), 59-69.

[22] Shin, S. (2021), A study to design the instructional contents for national curriculum of computer education in elementary school. *Journal of The Korean Association of Information Education*, 25(1), pp. 13-31.

[23] Shin, S. (2022), A study on educational data mining for public data portal through topic modeling method with latent dirichlet allocation. *Journal of The Korean Association of Information Education* 26(5), 439-448.

[24] Shin, S., & Bae, Y. (2015), A study on the hierarchical instructional system design of software education by school system, *Journal of The Korean Association of Information Education*, 19(4), 533-544.

[25] Stodden, V. (2020), The data science life cycle:

a disciplined approach to advancing data science as a science. *Communications of the ACM*, 63(7), 58-66.

[26] Wing, J. M. (2019), The data life cycle. *Harvard Data Science Review*, 1(1), 6.

저자소개



신 승 기

2017 University of Georgia(Ph.D.)
 2016~2017 미국 칼빈슨 정부연구소
 연구원
 2019~2020 에리조나주립대학교
 컴퓨터교육전공 교수
 2020~현재 서울교육대학교
 컴퓨터교육과 교수
 관심분야: Computational Thinking,
 인공지능교육
 e-mail: skshin@snue.ac.kr