

비전공자 대상의 데이터과학 교과목 개발과 컴퓨팅 사고력 효과 분석

Development of Data Science Course and Analysis of Computational Thinking Effect for non-Major Students

김재경[†] · 손의성^{††}

Jae-Kyung Kim[†] · Eui-Sung Sohn^{††}

요 약

데이터과학은 데이터를 통하여 자연과 사회 현상을 이해하고 의미 있는 정보와 지식을 발견하여 문제 해결을 하는 과학적 방법론이다. 데이터를 수집하여 분석하고 알고리즘으로 결과를 도출하는 데이터과학의 문제 해결과정은 올바른 컴퓨팅 사고를 필요로 하며, 다양한 전공 분야에 융합되어 실용적인 활용이 가능한 학문이다. 본 논문에서는 학습자의 컴퓨팅 사고력과 효능감 향상을 위해 비전공자를 대상으로 하는 데이터과학 교양 과목을 ADDIE 모델에 따라 개발하고 교육 효과를 평가하였다. 이를 위하여 컴퓨팅 사고 요소에 맞게 데이터과학의 문제 해결 과정을 단계별로 설계하였으며, 콘텐츠 개발 및 교육을 실시하고 평가 단계에서 사전·사후 검사를 통해 교육 효과를 분석하였다. 제시된 컴퓨팅 사고력 측정 도구를 이용하여 Likert 5단계 척도로 조사한 결과, 학습자의 컴퓨팅 사고력이 평균 3.12점에서 3.97점으로 향상된 것으로 나타났으며, 효능감 및 SW에 대한 흥미도 3.78점에서 4.22점으로 향상되었음을 확인하였다($p < 0.05$).

주제어: 데이터과학, 컴퓨팅 사고, ADDIE, SW교육, 비전공자 교육

ABSTRACT

Data science is a scientific methodology that understands actual phenomena through data and solves problems by discovering meaningful information and knowledge. The problem solving process of data science, which collects and analyzes data and derives results with algorithm, requires clear computational thinking and is a discipline that can be used practically by being fused to various major fields. In this paper, in order to improve learner's computing thinking ability and sense of efficacy, data science liberal arts course for non-majors were developed according to the ADDIE model. The data science problem-solving process was designed according to the corresponding computational thinking elements and education were conducted to improve the corresponding competency. The education effects were analyzed through pre- and post-tests. As a result, using the proposed CT evaluation tool of 5 point Likert scale, the learner's computing thinking ability was improved from 3.12 to 3.97 points on average, and the sense of efficacy and interest in SW improved from 3.78 to 4.22 points ($p < 0.05$).

Keywords: Data science, Computational thinking, ADDIE, SW education, Non-major undergraduates

1. 서론

오늘날 사회적, 경제적 및 자연적 이슈는 방대한 디지털 데이터로 생산되고 있으며, 이를 가공하여 새

운 지식을 창출하는 데이터 과학은 지능·정보산업의 새로운 패러다임이 되었다[1].

데이터과학은 실세계의 다양한 데이터를 수집 및 분석하여 유의미한 결과를 도출하는 과정으로, 우리

[†]정회원: 연세대학교 학부대학 조교수(교신저자)

^{††}정회원: 연세대학교 학부대학 조교수

논문투고: 2021년 02월 13일, 심사완료: 2021년 03월 29일, 게재확정: 2021년 04월 23일

* This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (2019R1G1A1010839).

가 사회 현상을 정확하게 이해하고 최적의 의사결정을 내릴 수 있도록 하는 학문 분야이다. 컴퓨터과학 및 통계학과 다양한 전공 지식을 융합한 데이터과학은 정보사회에서 학생들이 갖추어야 할 필수 기초 역량이다[2].

미국의 경우, 국립학술원에서 데이터과학 교육 강화 방안을 2018년 5월에 발표하고 대학에서 데이터과학자를 양성하기 위해 노력하고 있다. 미국은 141개 학교에서 데이터과학 프로그램을 개설하고 있으며 ‘데이터과학 기초’와 같은 교양 수업부터 컴퓨터과학, 통계학, 경영학과의 연계를 통한 데이터과학 전공을 신설하고 있다. 이외에도 데이터과학 관련 인증서 제도, 온라인 공개강좌, 부트 캠프 등 다양한 프로그램을 제공하고 있다[3].

또한 영국은 149개, 독일 22개, 캐나다 20개, 그리고 호주는 19개의 데이터과학 석사과정 프로그램을 운영하고 있다. 영국은 적극적으로 데이터과학 교육에 앞장서고 있고 독일은 데이터과학을 이용한 스마트 공장 및 로봇 분야에서 최첨단 기술을 보유하고 있다[4].

우리나라에서도 정부 부처 및 과학기술정보통신부의 주도로 ‘인공지능(AI) 국가전략’을 수립하고 2019년 12월 발표하여 전 국민이 AI 기초 역량을 습득할 수 있는 교육체계를 구축하는 것을 목표로 하고 있다. 이를 위해 AI대학원 프로그램 확대·다양화하고 교원의 양성·임용과정부터 SW·AI 기초 과목 이수 필수화하기로 하였다[5].

국내 대학의 경우 그동안 컴퓨팅 사고력 및 프로그래밍 교육이 활발히 이루어져 왔다. 최근에는 SW중심대학 2단계 사업[6]과 함께 인공지능 및 빅데이터 교육을 확대하며 새로운 지능정보기술 분야의 수요에 대응하고 있다.

Mckinsey Golbal Institute 보고서[7]에 따르면 미국의 경우 데이터과학 분야의 인력 수요는 2018년까지 14만~19만 명이 부족할 것으로 예상하였다. 국내의 경우도 한국데이터진흥원의 2015 데이터산업 현황 조사에[8] 따르면 업계에서 필요한 데이터 인력은 21,333명으로 기존 인력 대비 21.2%가 부족한 것으로 나타났다.

이에 대학에서 다양한 데이터과학 전공이 신설되고 있으며 생비전학습자의 컴퓨팅 사고력을 이용하여 학습자의 전공 영역의 다양한 데이터 처리를 통한 문제 해결 과정으로 이어지기 위해서는 데이터과학 교육이 활성화되어야 한다.

본 논문에서는 대학교 신입생 비전공자들을 대상으

로 한 데이터과학 교양 과목의 커리큘럼을 설계하였다. 제안 커리큘럼은 Python 언어와 가상환경 기반 패키지 관리 프로그램인 Anaconda를 사용한 프로그래밍 실습 위주의 내용으로 구성하였다. 또한 컴퓨팅 사고에 의한 문제 해결 과정에 맞춰 문제 정의, 데이터 수집 및 분석, 시각화, 기계학습 알고리즘 적용과 결과 도출까지 데이터 기반 문제 해결 과정을 체계적으로 학습할 수 있도록 하였다.

실험에서는 개발된 과목의 교육 효과를 측정하기 위해 데이터과학 교육을 통한 컴퓨팅 사고력의 강화와 실용적인 컴퓨팅 문제해결을 통한 학습자의 효능감 향상을 분석하였다.

2. 이론적 배경

2.1 데이터과학

데이터과학이라는 용어는 2001년도에 Cleveland에 의해 처음 사용되었는데[9], Cleveland는 통계학의 영역에 컴퓨터과학 등의 다양한 전공 분야를 확장한 학제 간 학문 분야의 필요성을 주장하며 이를 데이터과학 분야로 설명하였다.

이후 Drew Conway[10]는 Figure 1과 같은 데이터과학의 구성요소를 벤 다이어그램으로 나타냈으며, 전공지식에 해당하는 실질적인 전문지식(Substantive Expertise), 통계학에 해당하는 수학 및 통계 지식(Math & Statistics Knowledge), 그리고 컴퓨터과학에 해당하는 해킹 기술(Hacking Skills)의 교집합으로 표현하였다.

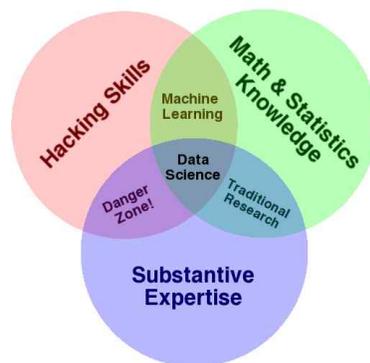


Figure 1. Data Science Venn Diagram

특히 올바른 수학 및 통계학 지식이 없는 상태에서 컴퓨터과학 기술과 전공지식만을 가지고 문제를 해결하는 것은 위험하므로 이 부분을 위험 지역(Danger

Zone!)으로 정의하며 모든 영역의 지식을 균형 있게 사용해야 함을 강조하였다.

데이터과학의 문제해결 과정은 Figure 2와 같이 연구자에 따라 다양한 모델이 사용되고 있으나 기본적으로 주어진 문제를 이해하고 필요한 자료를 수집하여 분석에 적합한 형태로 가공한 뒤, 이를 분석하여 문제 해결 과정을 구현하여 결과를 활용하는 단계로 이루어져 있다.

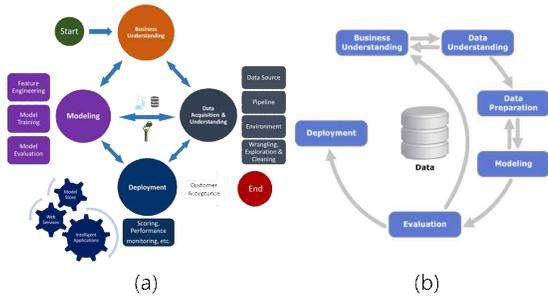


Figure 2. Example of Data Science Processes by (a) Microsoft (b) IBM

2.2 선행 연구

데이터과학에 기반한 문제해결 방식은 학습자의 컴퓨팅 사고력을 강화하고, 문제해결 과정에서 다양한 전공지식을 활용하는 실용성이 높은 교육 방법이다. 그러나 국내에서 비전공자를 대상으로 한 데이터과학 교육에 관한 연구는 아직 초기 단계이다.

허경은 학부생을 대상으로 엑셀을 이용한 데이터과학 실습 커리큘럼을 개발하고 학습자들의 만족도를 조사하였다. 제안 과목은 데이터를 가공하는 과정에 집중할 수 있도록 설계되었으며 Figure 2.(b)의 기계학습 단계인 모델링은 학습 내용에서 제외되었다[11].

김용민 및 김중훈도 엑셀을 활용한 데이터 과학 교육 프로그램을 개발하고 6일간 4~6학년의 초등학생을 대상으로 교육을 시행하였다. ‘엑셀의 기본 기능 익히기’, ‘스스로 데이터를 수집하여 나에게 안전사고가 자주 일어나는 이유 알아보기’ 등의 활동을 제안하고, 교육 후 컴퓨팅 사고력 향상 효과를 측정하였다[12].

김용민은 또한 애플리케이션을 이용하여 초등학생을 대상으로 PBL 기반 데이터과학 프로그램을 운영하여 ‘초등학생의 평균 수면시간과 키 성장의 관련’ 과 같은 문제를 적용하고 컴퓨팅 사고력 및 창의성 변화를 분석하였다[13].

선행 연구와 달리 본 논문에서는 비전공 대학생들이 데이터과학을 실용적으로 활용할 수 있도록 Anaconda 패키지 관리 프로그램과 범용 프로그래밍 언어인 파이썬 개발 환경에서 컴퓨팅 사고력 요소에 따른 데이터과학 교육 과정을 개발하고 교육 효과를 측정하였다.

컴퓨팅 사고력 요소는 문헌마다 다소의 차이는 있으나 일반적으로 ISTE (International Society for Technology in Education)와 CSTA(Computer Science Teachers Association)에서 제안하는 요소 중 데이터 수집, 데이터분석, 데이터표현, 문제분할, 추상화, 알고리즘 및 자동화 역량을 측정하였다.

3. 데이터과학 과목 개발 과정

본 논문에서는 교과목 개발에서 보편적으로 널리 사용되는 ADDIE 모델[14]에 따라 데이터과학 교과목을 개발하였다. ADDIE 모델은 Figure 3과 같이 분석, 설계, 개발, 실행, 그리고 평가의 5단계로 이루어져 있으며 요구분석에 의한 명확한 설계 및 개발이 이루어지고 단계별로 피드백을 반영할 수 있어 교과목 개발에 널리 사용되고 있다.

먼저 분석 단계에서는 학습자의 요구와 교육을 통하여 기대되는 역량 향상 등을 분석하여 개발 내용이 학습자의 인지적 및 정의적 영역의 발달에 적합한가를 확인한다.

설계는 학습 목표와 교수 전략 및 매체 설정에서 적용할 준거들로 이루어진다. 데이터과학의 문제해결 과정에 따른 단계적 내용의 설계, 컴퓨팅 사고력 향상을 위한 내용 및 실습 매체의 적합성 등을 확인한다.

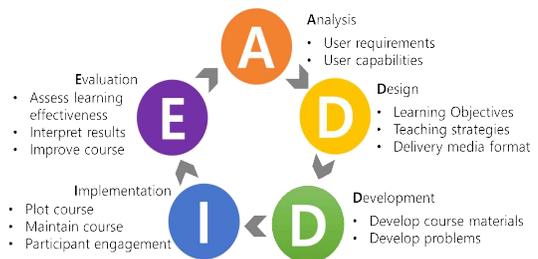


Figure 3. The Five Phase of ADDIE Model

개발 단계는 학습 목표와 설계에 따른 내용 구현의 적절성과 연관되므로 수업 자료, 연습 문제 및 평가 항목을 개발하고 검토한다.

실행 단계에서는 과목 개설, 유지보수, 그리고 환류

체계 등의 항목을 시행하여 수업을 운영한다. 마지막으로 평가 단계에서는 기대한 교육적 효과를 측정하기 위해 평가 대상과 적절한 평가도구를 선정한다. 측정된 결과를 분석하여 기대 효과의 달성 여부와 보완할 부분을 분석하여 과목 개발 단계를 다시 수행한다.

3.1 분석 단계

분석 단계의 사용자 요구 분석에서는 2019년 2학기 컴퓨팅 사고 및 프로그래밍 교양 수업을 수강하는 학생들을 대상으로 향후 ICT 교양 과목을 추가로 수강할 경우 희망하는 주제와 학습을 원하는 이유를 설문 조사하였다.

수강생 총 145명의 답변을 받아 분석한 결과는 Figure 4와 같다. 학습을 희망하는 주제로 인공지능 62명 (42.8%), 데이터과학 58명(40.0%), 모바일 컴퓨팅 8명 (5.5%), 기타 17명 (11.7%, 오피스 문서 및 멀티미디어 도구 등)의 응답을 하였다. 4차 산업에서 가장 인기 있는 분야인 인공지능에 대한 수요가 가장 많았으며 데이터과학도 비슷한 수요가 있었다. 인공지능 과목은 현재 개발 및 운영이 이루어진 상태이며 데이터과학 과목의 높은 수요와 개발의 필요성을 확인하였다.

과목 수강 희망 이유를 설문한 결과는 추가 SW 학습을 통하여 실용적인 컴퓨팅 문제해결 능력의 습득 기대 59명 (40.7%), 다양한 SW 분야 학습에 대한 흥미 40명 (27.6%), 프로그래밍 자신감 얻기 37명 (25.5%), 기타 9명 (6.2%, 복수전공 준비, 산업 트렌드) 순으로 나타났다. 따라서 이와 같은 학습자의 요구사항을 만족하도록 설계 방향을 수립하도록 한다.

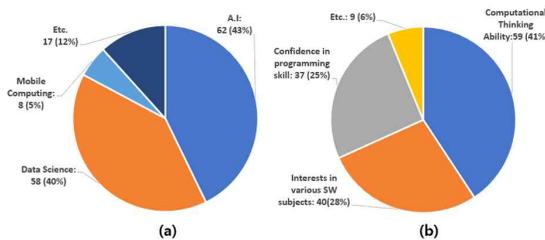


Figure 4. Students Survey Results

(a) Desired Topics (b) Reasons for Taking the Course

3.2 설계 단계

데이터과학은 당면한 문제를 해결하기 위해 다양한 데이터로부터 의미 있는 정보와 지식을 창출하여 최

적의 의사결정을 내리는 문제해결 과정이며 이것을 학습 목표로 설정하였다. 데이터과학의 문제해결 과정은 문헌마다 다소 상이한데 공통적인 핵심 단계를 정리하고 각 단계에 대응하는 주요 컴퓨팅 사고력 학습 개념을 Figure 5와 같이 정의하여 수업 내용을 설계하였다.

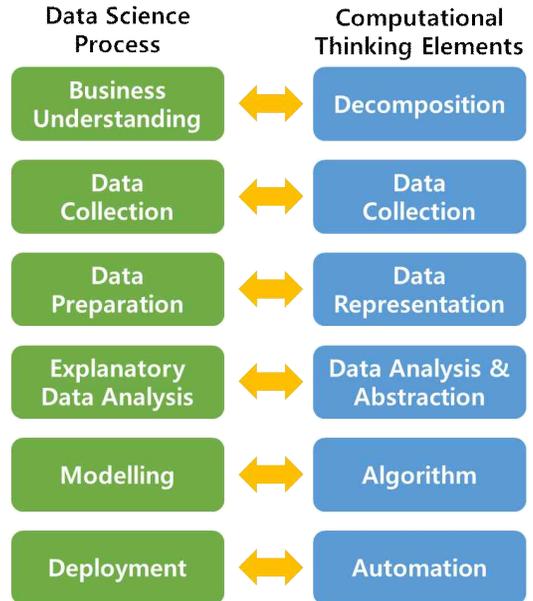


Figure 5. Design of Data Science Process and Corresponding Computational Thinking Elements

비즈니스 이해는 주어진 상황을 이해하고 도메인 지식을 이용하여 문제를 분석하는 단계로 컴퓨팅 사고력 문제 정의 범주의 분할 요소에 해당한다. 요구사항 정의, 필요한 데이터 정의, 문제해결의 성공기준 정의하기, 문제를 분할하여 단계별로 해결하는 전략을 학습한다.

데이터 수집 및 준비 과정은 이전 단계에서 정의한 데이터를 어디에서, 어떻게, 어떤 형식으로 수집할 것 인지를 학습한다. 데이터를 수집할 수 있는 소스와 수집 방법, 그리고 다양한 데이터 형식을 학습한다. 이 과정에서는 컴퓨팅 사고 요소의 데이터 수집 능력의 향상을 위해 문제해결에 필요한 실험 자료의 관찰 및 기록, 데이터 포털을 통한 검색, Open API를 이용한 실시간 자료 수집 방법을 학습한다.

데이터 준비 단계에서는 데이터를 분석에 적합한 정형화된 표 구조로 변환하고 기초적인 다듬기 작업을 학습한다. 컴퓨팅 사고의 자료 표현 방법에 따라

데이터를 표와 그래프로 구조화시켜 표현하는 것을 학습 목표로 한다.

데이터 탐색에서는 탐색적 데이터 분석(Explanatory Data Analysis) 학습이 이루어지며 통계 함수, 상관관계 분석, 데이터 시각화를 통하여 데이터를 이해하고 숨겨진 패턴과 필요한 특성값들을 찾아내는 과정을 학습한다. 이 과정에서는 데이터를 이해하고 일반화된 패턴을 찾아 가설을 수립할 수 있는 컴퓨팅 사고의 데이터 분석 능력과 세부 자료에서 필요한 정보만을 표현하는 추상화 개념을 익히도록 한다.

모델링 과정에서는 기계학습을 통하여 일반화된 현상을 예측하도록 하며 다양한 기계학습 알고리즘의 원리와 기능을 이해하고 문제해결에 적합한 알고리즘을 적용할 수 있도록 한다. 이는 문제의 해결 과정을 올바른 순서의 단계적인 작업들로 구성하는 컴퓨팅 사고의 알고리즘 사고 절차에 해당하며 필요한 일련의 연산들을 작성하는 능력을 키우도록 한다.

마지막으로 배포 단계는 컴퓨팅 사고의 자동화 단계로서 지금까지 진행한 데이터과학 문제해결 과정을 프로그래밍 명령어로 표현하여 실행을 자동화하고 비즈니스 이해 단계에서 정의한 요구사항 및 성공적인 해결 기준의 만족 여부를 평가하는 능력을 키운다.

교수 전략 및 매체 선정에서는 범용적인 문제해결 과정을 구현하기 위해서 파이썬 프로그래밍 언어와 Anaconda를 선정하였다.

3.3 개발 단계

개발 단계에서는 분석 단계의 학습 목표와 설계 내용에 따라 교수 자료, 연습 문제 및 평가 항목을 개발하였다.

학습자 분석에 따르면 학습 대상자는 대학 1학년생으로 프로그래밍 경험이 없거나 초·중등 교육 과정에서 기초 내용을 학습한 입문자 수준의 비전공생들이므로 1~3주차는 데이터과학의 소개 및 파이썬 프로그래밍 학습으로 구성하였다.

4주차에는 문제 정의 및 명세 작성하기와 데이터 수집하기 내용으로 구성하였다. Netflix Prize의 데이터과학 문제 정의 사례로부터 문제의 목표, 범위, 성공기준, 제약 조건, 수집할 데이터를 파악하고 데이터 과학 해결 절차에 따라 문제를 분할하여 해결하는 절차를 학습한다.

데이터 수집 학습에서는 웹 스크래핑, Kaggle, 공공 데이터포털 Open API 활용하기와 같은 기술적인 테

이터 수집 방법과 XML, REST, CSV와 같은 주요 데이터 포맷을 학습하고, 실생활 문제를 통하여 지하철 및 버스노선, 미세먼지, COVID-19 정보 등 실용 데이터를 수집하는 연습을 한다.

예시로 Figure 6은 지하철역 이름을 입력받아 역 주변의 버스 번호를 실시간으로 공공데이터포털의 API로 수집하여 출력하는 문제로 문제해결을 위해 필요한 데이터의 종류, 수집 위치, 데이터 형식, 수집 방법을 학습자 스스로 해결하는 사례이다.

In-Class Problem: Enter the name of a subway station from the user and print out a list of bus numbers available around the station.

- What data do you need to solve the problem?
- How do you get bus and subway station data?
- Which format is the data collected?
- What is the expected and correct output?

User Input

```
station_name = input("지하철역 이름: ")
```

지하철역 이름: 캠퍼스타운

Expected Output

```
['16', '35', '51-1', '6-3', '91', '92', '9201']
```

Figure 6. Example of Data Collection Problem

5주차에는 정형 데이터의 이해와 표 데이터의 행렬 추가, 삭제, 변환 등과 같은 기본적인 연산을 학습하여 수집한 데이터를 표 형식으로 표현한다. 표 구조의 데이터 표현과 숫자, 문자열, 시계열 데이터 등 다양한 자료 유형을 능숙하게 다룰 수 있도록 실습 위주의 수업으로 구성한다.

6~7주 탐색적 데이터 분석에서는 기초 통계 함수와 피어슨 상관계수를 이용하여 데이터의 요약 정보와 특성 간의 관계를 파악하는 방법을 학습하고, 수치 및 범주 데이터 간의 시각화를 통하여 정보를 추상화하여 파악하는 훈련을 한다. 수치 데이터 간의 시각화 방법인 산점도, 수치-범주 데이터의 막대 및 상자수염 도표, 범주 데이터 간의 히트맵과 모자이크 플롯을 통해 데이터에 숨겨진 패턴을 추상화된 시각적 정보로 파악할 수 있는 능력을 익힌다.

Figure 7은 7주차 연습 문제의 일부분으로 음식점에서 성별에 따라 봉사료 지불 금액의 차이가 있는지 파악하기 위해 실제 음식점 매출 데이터를 수집하고 세부 데이터를 선형회귀 도표로 추상화하여 문제해결에 필요한 패턴을 찾는 예시이다.

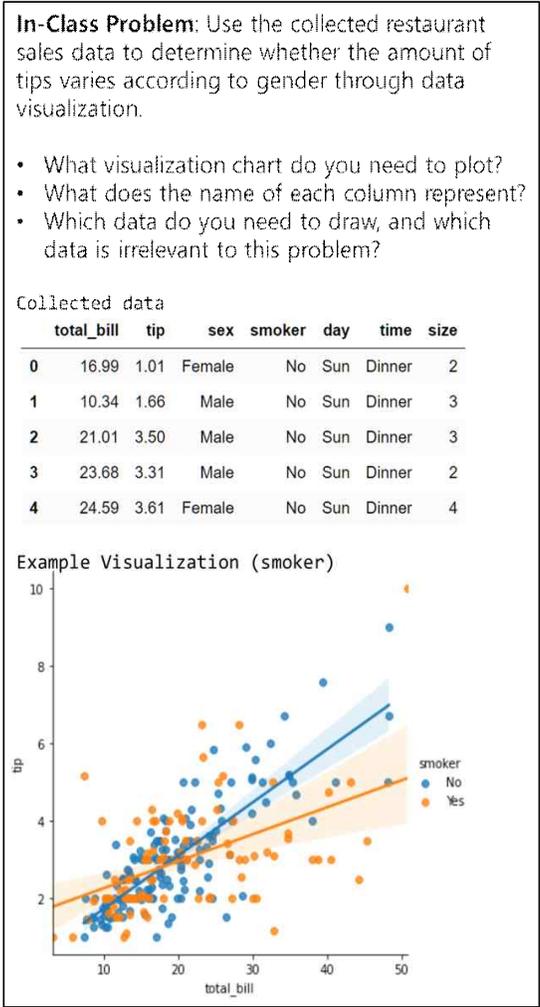


Figure 7. Example of Data Visualization Problem

중간 평가 이후 9주차에는 기계학습 알고리즘의 원리와 종류를 소개하고, 10주부터 13주까지 다양한 알고리즘을 학습하고 수집한 데이터를 이용하여 예측하여 본다. 선형회귀 분석, SVM, k-근접이웃, 의사결정나무, 랜덤 포레스트, 나이브 베이즈, 신경망 알고리즘의 원리를 이해하고 주어진 데이터의 특성에 따라 최적의 알고리즘을 적용하여 해를 구하도록 한다.

14주에는 컴퓨팅 사고의 자동화 요소인 프로그램의 정확성과 오류 확인을 학습하기 위해 평가 매트릭스인 ASE, MSE, RMSE, ROC 학습하고 예측 결과의 확인 및 오류 수정 방법을 실습하여 문제해결 과정을 완성하도록 한다. 마지막으로 해결 과정을 SW 구현물로 완성하여 데이터 입력에 대해 자동화된 예측 결과를 도출하는 모델을 배포하는 과정을 실습한다.

3.4 실행 단계

개발된 과목은 대학교 1학년을 대상으로 하는 IT 교양으로 개설하였으며 주 2시간의 이론 수업과 1시간의 문제 풀이 실습 시간으로 총 3시간씩 15주간 운영하였으며 강의 계획서는 [15]에서 확인할 수 있다.

강의는 실시간 영상 수업으로 진행하고 수업 후 녹화된 수업 영상을 학습자들이 언제든지 시청할 수 있도록 하였다. 성취도 평가는 매주 부여되는 실습 결과물과 중간시험 및 학기 말 프로젝트로 이루어졌다. 또한, 중간시험 이후 수강생 의견을 수렴하여 문제 난이도 조정을 시행하였다.

4. 평가

4.1 연구 대상

과목 개발의 평가 단계에서는 분석 단계에서 기대한 교육적 효과가 나타나는지를 평가하였다. 이를 위해 대학교 신입생을 대상으로 2020년 2학기에 개설한 15주 기간의 정규 데이터과학 입문 교양 수업에서 수강생들의 컴퓨팅 사고력 핵심 요소의 향상과 컴퓨팅 사고력 효능감 및 흥미 향상에 효과가 있는지를 측정하였다.

수강생은 모두 신입생으로 컴퓨터과학 및 데이터과학 비전공자였으며 이학·공학계열 6명 (16.2%), 인문·사회계열 25명 (67.6%), 의예 6명 (16.2%)의 총 37명으로 구성되었다.

4.2 연구 도구

현재 표준화된 컴퓨팅 사고력 검사 도구는 존재하지 않으나 이와 관련한 다양한 연구들이 진행된 상태이다[16,17]. 본 논문에서는 컴퓨팅 사고력 핵심 요소를 평가하기 위해 이민우[18]의 범용 컴퓨팅 사고력 척도 도구의 문제 이해, 추상화, 알고리즘 절차, 자동화 영역의 20개의 자기 평가 설문을 본 수업의 프로그래밍 언어와 문제 해결환경에 맞게 수정하여 Likert 5 단계 척도로 사용하였다.

또한 컴퓨팅 사고에 대한 효능감 및 흥미 향상을 측정하기 위해 박주연[19]이 개발한 정의적 영역의 13개 세부 평가 요소 중 본 수업의 자신감과 학습 흥미의 2개 영역의 5개 문항을 동일한 척도의 자기 평가 설문으로 사용하였다. 최종 설문 영역과 문항에 대한 Cronbach α 값은 Table 1과 같다.

Table 1. Evaluation items by CT Elements

Category	Elements	No.	Cronbach α
CT Capability	Decomposition	3	0.87
	Data Collection	3	0.90
	Data Representation	2	0.84
	Data Analysis	3	0.89
	Abstraction	3	0.91
	Algorithm	3	0.83
	Automation	3	0.79
Self Efficiency	Confidence	2	0.86
	Interest in SW	3	0.82

4.3 연구 결과

개발된 수업의 효과를 검증하기 위해 강의 시작 시점의 사전 평가와 강의 종료 시점의 사후 평가를 결과로 Shapiro-Wilks 검정을 시행하여 모두 정규성을 가지는 것을 확인한 후, 대응표본 T-검정을 시행하였다.

Table 2. Effectiveness of CT capability result before and after the lecture.

CT Element	Type	Mean	S.D	t
Data Collection	Pre	3.06	0.87	-5.38*
	Post	3.92	0.79	
Data Representation	Pre	3.00	1.01	-6.53*
	Post	4.18	0.72	
Data Analysis	Pre	2.64	0.98	-9.19*
	Post	4.34	0.67	
Decomposition	Pre	3.32	1.08	-2.56**
	Post	3.78	0.99	
Abstraction	Pre	3.65	0.52	-8.86*
	Post	4.25	0.46	
Algorithm	Pre	2.83	0.65	-4.09*
	Post	3.61	0.71	
Automation	Pre	3.36	0.87	-2.37**
	Post	3.70	0.79	

* $p < 0.01$, ** $p < 0.05$

먼저 컴퓨팅 사고력 핵심 요소의 경우 Table 2와 같이 5개 영역 모두 사후 점수의 평균이 사전 검사 평균 점수에 비해 높게 측정되었으며 모든 요소가 유의 수준 1%와 5%에서 검정하였을 때 유의미성을 나타내었다. 특히 데이터의 수집, 표현, 분석과 추상화 부분은

다른 요소에 비해 상대적으로 높은 교육 효과가 있었음을 알 수 있었다. 결과적으로 개발된 데이터과학 교육을 통하여 컴퓨팅 사고력을 이용한 문제해결력이 향상된 것으로 확인되었다.

다음으로 효능감 측정에서 자신감과 학습 흥미도에 대한 사후 점수의 평균이 Table 3과 같이 유의미하게 높음을 보였다. 이로써 개발된 교육 과정을 통하여 학습자가 컴퓨팅 사고력 효능감을 고취시키고 SW영역에 대한 흥미도를 높일 수 있음을 확인하였다.

Table 3. CT self-efficiency result before and after the lecture.

Self Efficiency	Type	Mean	S.D	t
CT Confidence	Pre	3.71	1.08	-3.15*
	Post	4.38	0.75	
SW Interests	Pre	3.84	0.47	-2.24**
	Post	4.07	0.53	

* $p < 0.01$, ** $p < 0.05$

본 연구의 컴퓨팅 사고력 측정과 별개로 수업 종료 후 추가로 시행한 강의평가 주관식 설문을 살펴보면 일반적인 프로그래밍이나 컴퓨팅 사고 입문 과목에서는 배울 수 없는 실용적인 데이터 처리 기술을 익힘으로써 대다수 학생들은 높은 자신감과 향후 SW관련 학습에 대한 흥미를 얻은 것으로 분석되었다. 반면 기초 컴퓨팅 과목에 비해 다소 난이도가 높은 내용으로 인해 2명의 학생은 흥미가 사전 검사에 비해 낮아지는 사례도 있었다. 이는 향후 과목의 유지보수에서 보완해야 할 사항이다. 또한, 수강생 대부분이 사전 검사 시점에 이미 SW 분야에 높은 흥미를 느끼고 있어 평균의 차이는 크지 않았다.

5. 결론

본 논문에서는 대학교 신입생을 대상으로 컴퓨팅 사고력을 향상시키고 SW분야의 흥미를 높이기 위한 비전공생을 위한 데이터과학 교과목을 개발하였다.

개발 과목은 ADDIE 모델에 따라 학습자의 요구를 분석하여 컴퓨팅 사고력의 향상과 자신감 및 흥미를 고취를 목표로 설계를 하였다. 이를 위해 데이터과학의 비즈니스 이해, 데이터 수집 및 준비, 탐색적 데이터 분석, 모델링, 배포 단계를 컴퓨팅 사고의 문제 정의, 데이터 수집, 분석 및 표현, 추상화, 알고리즘, 자동화 요소 개념을 적용하여 설계하고 해당 역량을 향

상시킬 수 있는 내용을 설계하였다.

개발된 내용은 대학교 비전공 신입생을 대상으로 한 교양 교과목으로 개설하여 15주간 교육하였으며, 평가 단계에서는 컴퓨팅 사고력의 향상과 자신감 및 흥미 향상의 효과성을 분석하기 위해 사전·사후 자기 평가 설문을 시행하였다. 그 결과 개발된 데이터과학 교육을 통하여 학습자의 컴퓨팅 사고력을 향상시켰으며 문제해결의 자신감과 SW분야에 대한 흥미도 늘어난 것으로 나타났다.

향후에는 데이터과학 교육을 통한 컴퓨팅 사고력 뿐 아니라 데이터활용 역량 향상을 평가할 수 있는 도구가 개발되어야 할 것이다.

참고문헌

- [1] Hey, T., Tansley, S., & Tolle, K. (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Washington. Microsoft Research. DOI: 10.1007/978-3-642-33299-9_1
- [2] Raj, R. K., & Parrish, A. (2019, July). Data Science Education: Global Perspectives and Convergence. *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*. (pp. 265-266). NY, USA. DOI: 10.1145/3304221.3325533
- [3] National Academies of Sciences, Engineering, and Medicine. (2018). *Data Science for Undergraduates: Opportunities and Options*. Washington DC. The National Academies Press.
- [4] Park, S-H. (2019). A Study on Data Scientist Training Strategies for Preparation of th 4th Industrial Revolution Era. *The National Academy of Sciences*, 58(2). 307-400.
- [5] Ministry of Science and ICT. (2019, Dec 17). *AI National Strategy Announcement*. Ministry of Science and IC T. <https://www.korea.kr/news/pressReleaseView.do?newsId=156366736>
- [6] Lee, J. (2020, September 3). *Ministry of Science and ICT confirms software-centered university phase 2 project*. eToday News. <https://www.etoday.co.kr/news/view/1936795>
- [7] Manyika, J., Brown, B., & Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. Mckinsey Golbal Institute.
- [8] Korea Database Agency. (2015). *Data Industry Status Survey Report*. Seoul. Korea Database Agency.
- [9] Cleveland, W. S. (2001). Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review*, 69(1), 21-26, DOI:10.2307/1403527
- [10] Conway, D. (2013, Mar 26). *The Data Science Venn Diagram*. Drew Conway Data Consulting. <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- [11] Hur, K. (2020). Curriculum of Basic Data Science Practices for Non-majors. *Journal of Practical Engineering Education*, 13(2), 265-273. DOI : 10.14702/JPEE.2020.265
- [12] Kim, Y., & Kim, J. (2017). Effect of data science education program using spreadsheet on improvement of elementary school computational thinking. *Journal of The Korean Association of Information Education*, 21(2), 219-230. DOI : 10.14352/jkaie.2017.21.2.219
- [13] 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), 551-562. DOI : 10.14352/jkaie.2020.24.6.551
- [14] ADDIE Model. (2013, Mar 26). *ADDIE Model*, <https://www.instructionaldesign.org/models/addie/>
- [15] Data Science Syllabus. (2020, Sep 1). *Data Science Syllabus*, <http://swclass.yonsei.ac.kr:2020/dev/web/syllabus.html>
- [16] Kim, M., Lee, W., & Kim, J. (2017). Presenting the Development Direction Through the Analysis of Tool used to Measure Computational Thinking. *The Journal of Korean Association of Computer Education*, 20(6), 17-25.
- [17] Yang, J. (2017). *Software Education Competency Diagnosis and Analysis Research*. Daegu. Korea Education and Research Information Service.
- [18] Lee, M., & Kim, S. (2019). Study on the Development of a General-Purpose Computational Thinking Scale for Programming Education on Problem Solving. *The Journal of Korean Association of Computer Education*, 22(5), 66-77.
- [19] Park, J., Kim, J., Kim, S., Lee, H., & Kim, S. (2017). Development of evaluation factors for SW education in elementary and secondary schools. *The Journal of Korean Association of Computer Education*, 20(6), 47-59.



김재경

2007년 연세대학교 컴퓨터과학과
(공학박사)

2009년 연세대학교 컴퓨터과학과
연구교수

2014년 연세대학교 컴퓨터과학과
객원교수

2017년 ~ 현재 연세대학교 학부대학 조교수
관심분야: SW교육, 데이터과학



손의성

2012년 연세대학교 컴퓨터과학과
(공학박사)

2012년 ~ 2014년 연세대학교
소프트웨어응용연구소 연구원

2015년 ~ 2018년 애니펜 수석연구원

2018년 ~ 현재 연세대학교 학부대학 조교수
관심분야: SW교육, 딥러닝, 컴퓨터그래픽스