데이터과학 교육과정에 대한 분석적 연구*

An Analysis of Data Science Curriculum in Korea

이 혜 원 (Hyewon Lee)**
한 승 희 (Seunghee Han)***

목 차

초 록

이 연구에서는 2019년 10월 현재 우리나라에 개설된 데이터과학 교육과정의 현황을 분석하기 위해 먼저, 데이터과학 분야의 교육과정을 분석한 기존 연구와 데이터과학 분야 전문가에게 요구되는 역량에 대한 분석을 진행하였고, 이를 바탕으로 우리나라에 개설된 80개의 교육과정과 2,041개의 교과목을 대상으로 학문 영역 특징 기반 분석, 데이터 전문가 역량 기반 분석과 교과목명 내용 분석을 실시하였다. 분석 결과, 우리나라에서의 데이터과학 전공 교육은 기술과 직업 실무적 관점보다는 학문적 접근을 바탕으로 한 연구 중심의 전문적 교육과정으로 자리 잡았으며, 통계적 분석 역량을 중심으로 많은 교과가 개설되었고, 정보기술, 통계학, 경영학을 중심으로 한 학제적 특성이 교육과정에 반영되었음을 확인하였다.

ABSTRACT

In this study, in order to analyze the current status of the data science curriculum in Korea as of October 2019, we conducted an analysis of the prior studies on the curriculum in the data science field and the competencies required for data professional. This study was conducted on 80 curricula and 2,041 courses, and analyzed from the following perspectives: 1) the analysis of the characteristics of data science domain, 2) the analysis of key competencies in data science, 3) the content analysis of the course titles. As a result, data science program in Korea has become a research-oriented professional curriculum based on an academic approach rather than a technical, vocational, and practitional view. In addition, it was confirmed that various courses were established with a focus on statistical analysis competency, and interdisciplinary characteristics based on information technology, statistics, and business administration were reflected in the curriculum.

키워드: 데이터과학, 교육과정, 교과목 분석, 내용 분석, 데이터 전문가 역량, 데이터과학 생명주기
Data Science, Curriculum, Curriculum Analysis, Content Analysis, Competency of Data Professional, Data Science Lifecycle

---

* 이 논문은 2019학년도 서울여자대학교 교내학술연구비의 지원을 받았음.
** 서울여자대학교 사회과학대학 문헌정보학과 부교수(hwlee@swu.ac.kr / ISNI 0000 0004 6819 5799) (제1저자)
*** 서울여자대학교 사회과학대학 문헌정보학과 교수(hanshee@swu.ac.kr / ISNI 0000 0004 6824 6058) (교신저자)
논문접수일자: 2020년 2월 4일 최초심사일자: 2020년 2월 12일 게재확정일자: 2020년 2월 19일
한국문헌정보학회지, 54(1): 365-385, 2020. http://dx.doi.org/10.4275/KSLIS.2020.54.1.365

1. 연구의 목적 및 필요성

데이터과학(data science)이란 실제 현상을 이해하고 분석하기 위해 통계, 데이터 분석, 기계 학습 및 관련 방법을 통합하는 개념(Hayashi 1998)으로, 과학적 방법론과 알고리즘, 시스템 등을 사용하여 구조화되거나 구조화되지 않은 데이터로부터 지식과 통찰력을 추출하는 학제간 분야를 말한다(Dhar 2013). 즉, 분석하고자 하는 문제의 정의, 문제 해결을 위한 데이터의 수집・처리・분석 방법, 그리고 분석된 내용을 기반으로 한 추론 등 데이터를 활용한 과학적 예측의 모든 과정을 포함하며, 이를 위해 다양한 학문분야의 이론과 기법의 융합이 요구되는 학제적 성격의 학문분야라 할 수 있다.

지식관리 관점에서 조직의 효율적 운영을 위한 과학적 의사결정은 늘 관심의 대상이었고, 데이터의 효과적 수집과 분석은 이를 위한 주된 방법이 되어왔다. 최근 들어 정보기술이 발전하고 사람들의 다양한 활동들이 데이터로 축적되기 시작하면서 이러한 과학적 의사결정을 뒷받침해 줄 데이터의 규모가 커지고, 이로 인해 대규모 데이터를 과학적으로 처리할 수 있는 학문적 접근법이라 할 수 있는 데이터과학 영역이 주목받게 되었다. 그러므로 이 학문은, 어느 날 새롭게 출현한 것이 아닌, 기존의 다양한 학문들이 대규모 데이터에 대한 효율적 분석이라는 필요성에 의해 서로 융합하여 하나의 새로운 분야로 구축되었다고 할 수 있다.

4차 산업혁명의 핵심 기제인 빅데이터 관련 시장은 나날이 성장하고 있으며, 차세대 경제 성장 동력을 발굴하고 정보기술 중심의 경쟁구도에서 유리한 위치를 선점하기 위해 관련 전문인력을 양성하는 것은 전세계적으로 중요한 가치가 되었다. 우리나라에서는 2010년대 초반 대학원 과정을 중심으로 관련 전문인력을 양성하는 교육과정의 개설 및 운영을 시작한 이래로 빠른 양적 성장을 통해 2019년 현재 80개의 학부 및 대학원 과정에서 관련 인력을 양성하고 있다. 이러한 빠른 성장은 데이터 관련 전문인력 양성에 대한 사회적 요구를 반영한 것이라 할 수 있다.

이러한 양적 성장과는 별개로 교육의 내용적 측면은 어떻게 이루어지고 있는지 알아보기 위해, 이 연구에서는 2019년 현재 우리나라에 개설된 데이터과학 교육과정의 현황을 분석하고자 한다. 이를 위하여 먼저 선행연구에서는 데이터과학 분야의 교육과정을 분석한 기존 연구와 데이터과학 분야 전문가에게 요구되는 역량에 대한 분석을 진행하였고, 이를 바탕으로 우리나라에 개설된 80개의 교육과정을 대상으로 학문 영역 특징 기반 분석, 데이터 전문가 역량 기반 분석과 교과목명 내용 분석을 실시하였다.

2. 선행연구

2.1 데이터과학 전공 현황 및 교과목 분석

데이터과학 전공 현황 및 교과목을 분석한 연구는 국외와 국내로 구분하여 연도별로 정리하였다.

Gil(2014)은 컴퓨터 프로그래밍 관련 수업을 받지 않은 학생들에게 초점을 두어 데이터과학의 기본 개념을 교육하고 데이터 중심의 연구 프로젝트를 진행할 수 있는 방식 및 과정을 이해시

키고자 하였다. 또한 향후 그들이 컴퓨터 공학자들과 협력할 수 있는 능력을 배양시키고자 교육과정에 WINGS 워크플로우 시스템을 도입하였다. 워크플로우 시스템은 데이터 분석 뿐만 아니라 데이터 전처리, 후처리, 시각화에 이르기까지 데이터 전단계에 거쳐서 학생들이 처리해야 할 임무를 지정할 수 있으며 올바른 데이터 분석 방법을 설정했는지에 대한 검증도 제공하였다.

Lyon 등(2015)은 데이터 큐레이션의 업무를 주로 하는 데이터 사서, 데이터 관리자 및 큐레이터, 데이터 아키비스트 등의 채용에서 제시되는 직무분석표와 광고를 교육과정에 매핑하는 연구를 진행하였다. 그들은 데이터 사서, 데이터 관리자 및 큐레이터, 데이터 아키비스트 모두에게 필요한 역량으로 메타데이터 생성 및 관리 능력, 연구데이터 관리 및 구조에 대한 이해 능력, 특정 도메인에 대한 전문지식과 데이터에 대한 이해 능력, 통계분석 실행 능력 등을 제시하였다.

Zhang 등(2017)은 중국 내 데이터과학 교육과정을 비교하는 연구를 진행하였으며, 중국 내 데이터과학은 아직 시작 단계에 있으며 문헌정보학 학부과정부터 박사과정에 이르기까지 데이터 전문가로서의 역량을 갖춘 인재를 배출해야 한다고 주장하였다.

Thoma와 Urban(2018)은 도서관의 데이터 전문가를 배출하기 위해 문헌정보학 대학원의 교육과정에 변화가 필요하다고 주장하였다. 그들은 105명의 데이터 전문가에게 문헌정보학 대학원 교육과정이 본인의 업무를 수행하는데 얼마나 도움이 되었는지를 알아보았다. 또한 데이터 전문가의 성과를 상급관리자들이 어떻게 인식하고 있는지도 살펴보았다. 그들은 현장 경험을 포함하고 정보 환경에 맞게 변화된 MLIS 프로그램을 이수한 졸업생이 데이터 큐레이션 분야에 진출할 때 매우 유리하다는 것을 확인하였다. 특히 데이터 큐레이션 업무를 희망하는 사람이 데이터와 관련된 별다른 교육을 받지 않았다면 MLIS 프로그램이 크게 도움이 된다는 것을 확인하였다. 또한 iSchool과 LIS의 프로그램이 데이터 센터나 데이터과학 학위 프로그램과의 파트너십을 구축하여 몰입형 실무 교육(immersive embedded practical education)을 제공해야 한다고 주장하였다.

강지혜의 연구(2016a)에서는 69개 iSchool에서 데이터 사이언스 관련 교과가 어떻게 제공되고 있는지를 파악하고, 국내 교과와 비교하여 방향성을 제시하였다. 그는 iSchool이 건강, 기술, 바이오 분야를 비롯한 관련 분야로 그 교과 영역을 확장하는 것에 반해, 국내 교과에서는 인접학문과 융합하려는 현상이 활발하게 관찰되지 않았음을 확인하였다. 데이터를 어떻게 처리하고 관리할 것인지에 대한 영역 역시 iSchool이 집중하는 분야인데, 일반적인 데이터 사이언스, 데이터 관리, 데이터 보안 등에 중점을 둔 교과가 제공되고 있었다. 데이터를 저장하는 방식에 대한 교과 분류는 '데이터베이스' 관련 교과의 비중이 높았으며, 비슷한 비중으로 통계와 분석법이 제공되고 있었다(강지혜 2016b).

유사라(2018)의 연구에서는 정보 리터러시와 관련된 인지과정 중심 교육모형(Bloom's taxonomy), 사서 직무교육 내용, 디지털 큐레이션 등에 관한 최근 연구들을 바탕으로 메타리터러시 역량을 도출하였다. 웹3.0 정보환경의 상황적 요소를 분석할 수 있는 비판적 사고능력과 정보 신기술의 적응 능력을 확장시킨

메타리터러시 개념을 기준으로 문헌정보학 전공 대학 커리큘럼을 분석・비교하였으며, 미래 지향적 예비 사서의 전문성 배양 측면에서 강조되어야 할 전공 교과목을 진단하고 커리큘럼 개선을 위한 방안을 제시하였다.

이상철(2019)은 Conway나 빅데이터 참조모델의 핵심역량을 참고하여 교과과정 로드맵을 제안하였다. 그는 빅데이터과학의 핵심역량을 분석 역량, 기술 역량, 경영 역량으로 설정하여, 학부 1학년의 전공기초 교과목으로 기초수학, 통계학개론, 컴퓨터사고와 SW 코딩, 경영학개론 등을 제시하였다. 2학년의 전공기본 교과목은 R통계분석, 빅데이터 플랫폼(하둡), 마케팅원론 등이며, 3학년 전공심화 교과목은 파이션 기계학습, 빅데이터 시각화, 인사 및 재무관리 등이고, 마지막 4학년 전공실무 교과목은 빅데이터 애널리틱스, 빅데이터 캡스톤 디자인, 빅데이터와 윤리경영, 빅데이터 비즈니스모델 등이었다.

기존 연구들은 데이터과학의 교육과정 분석을 교과목 분석으로 그 범위를 축소하여 접근하였다. 데이터과학의 교과목 분석은 대학과 대학원을 구분하는 학문위계나 전체적인 학습 내용, 대학 구성원간의 협력, 현장학습을 다 포함하지는 못한다. 이 연구에서는 대학 및 대학원의 전체적인 교육과정을 분석하였다.

2.2 데이터과학 역량 분석

본 장에서 다룬 선행연구는 데이터 전문가에게 필요한 역량과 데이터과학 생명주기의 프로세스별 핵심 개념 분석으로 나누어 진행하였다.

먼저, 데이터 전문가에게 필요한 역량을 다룬 연구를 연대순으로 정리하였다.

이명호의 연구(2016)에서는 국내외 데이터 사이언티스트(Data Scientist) 양성을 위한 데이터 사이언스 프로그램의 교과과정을 분석하였다. 해당 연구의 분석 대상은 국내 7개 대학교와 미국의 10개 대학교로 삼았으며, 총 14개의 데이터 사이언스 과정이 대학원 중심으로 운영되고 있음을 확인하였다. 이명호의 연구(2016)에서는 Conway의 데이터 사이언스 영역으로 총 272개 과목을 분류하였으며, 그 결과 국내, 미국 모두 수학 및 통계 지식 영역의 교과목이 각각 10%, 26%로 가장 높은 비율을 보였다.

한국정보화진흥원(2014)은 데이터 과학자가 갖추어야 될 역량으로 기반 역량(Foundation Coverage), 기술 역량(Technology Coverage), 분석 역량(Analysis Coverage), 사업 역량(Business Coverage) 등 총 네 가지 영역으로 분류하였고, 해외 대학(원) 과정에서도 이러한 역량들을 바탕으로 교육과정이 구성되었다고 설명하였다.

Zawadzki(2014)는 데이터 과학자가 갖추어야 할 역량으로 수학 및 통계(Math & Stat), 프로그래밍 및 데이터베이스(Programming & Database), 도메인 지식과 소프트 스킬(Domain Knowledge & Soft Skills), 커뮤니케이션 및 시각화(Communication & Visualization) 등을 제시하고 각각의 세부사항을 정의하였다. 수학 및 통계는 기계학습, 통계 모델링, 클러스터링 등과 같은 데이터 분석에 초점을 두었고, 프로그래밍 및 데이터베이스는 Hadoop과 같은 프로그램 언어나 SQL, NoSQL 등과 같은 데이터베이스 처리 언어에 집중하였다. 도메인 지식과 소프트 스킬은 업무에 대한 열정이나 데이터 그 자체에 대한 호기심, 문제해결능력 등을 다루었다. 커뮤니케이션 및 시각화는 분석

결과에 대한 스토리텔링이나 시각적인 표현 기술에 집중한 것이었다.

PCMI(2016)는 데이터과학의 정의 및 주요 사항을 고려하여 데이터 기술 및 큐레이션(Data Description and Curation), 수학적 기초(Mathematical Foundations), 컴퓨팅 사고(Computational Thinking), 통계적 사고(Statistical Thinking), 데이터 모델링(Data modeling), 커뮤니케이션, 재현성, 윤리(Communication, Reproducibility and Ethics) 등 여섯 개의 데이터과학 역량을 정의하고 그에 맞는 커리큘럼을 완성하였다. 그들은 기본 영역인 수학에 기초를 두어 모든 과정에서 컴퓨팅 사고와 통계적 분석이 가능한 교육과정을 재설계하였다.

Tang과 Sae-Lim(2016)은 미국 대학의 데이터과학 프로그램에 대해 분석하였다. 그들의 연구에서는 8개 분야에서 무작위로 선정된 30개의 데이터과학 프로그램에 대한 탐색적 내용 분석을 실시하여 미국의 데이터과학 교육에 상당한 격차가 있음을 확인하였다. 분석 대상은 프로그램 설명, 커리큘럼 요구사항, 데이터과학 강의에 포함된 주요 기술, 도메인 지식 등이었다. 그들의 연구 결과에서 주목할 점은 iSchool의 주요 교과목이 커뮤니케이션과 시각화 기술을 강조하는데 반해, 대학의 데이터과학 프로그램은 그러한 기술들을 충분히 다루지 못한다는 것이다.

강지혜의 연구(2016b)에서는 국내 국가직무능력표준(National Competency Standards, NCS)의 직무분석을 적용한 데이터 사이언스의 교과목을 설계하기 위해 데이터 사이언스의 능력단위요소를 제안하였다. 해당 연구에서는 NCS 체계와 국외 연구에서 산발적으로 언급되는 데이터 사이언스 관련 '능력단위'를 통합하여 데이터 사이언스에 필요한 능력을 전반적으로 정리하였다.

시라큐스대학의 iSchool에서는 학제적인 성격이 강한 데이터과학의 교육과정을 네 개의 역량인 수집 및 조직(capture and organization), 기술적인 분석(technical analysis), 시각화 및 커뮤니케이션(visualization and communication), 실무 적용(practical application)으로 설계하였다(iSchool@Syracuse 2019).

<표 1>은 데이터과학 역량을 다룬 선행연구를 분석한 것이다. 이 연구에서는 선행연구에 나타난 역량을 6개 항목 '프로그래밍', '수학 및 통계; 분석', '도메인 지식', '커뮤니케이션 및 윤리', '데이터 수집, 처리 및 모델링', '시각화' 등으로 구분하고 이를 4.2장의 데이터과학 역량을 기준으로 한 교과목 분석의 기초자료로 활용하였다.

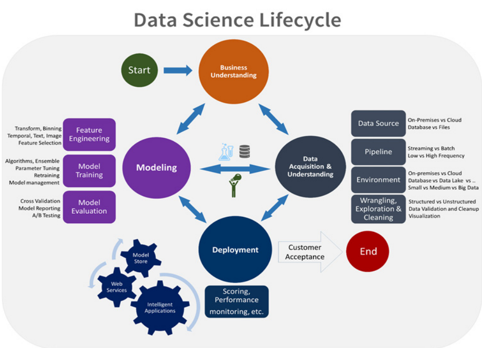
다음은 데이터과학 생명주기의 프로세스별 핵심 개념을 분석하여 데이터 전문가에게 필요한 역량을 도출한 선행연구이다.

The Team Data Science Process(TDSP)는 데이터과학 생명주기 프로세스(<그림 1> 참고)이며, 데이터과학 분야의 프로젝트를 구조화할 때 활용된다. TDSP의 데이터과학 생명주기 프로세스는 프로젝트를 진행하는 구성원들 간의 의사소통을 원활하게 할 뿐만 아니라 고객과의 커뮤니케이션 상에서도 오류가 발생하지 않도록 지원하여 성공적인 프로젝트를 보장한다. <그림 1>의 TDSP 데이터과학 생명주기 프로세스는 업무 이해(Business Understanding), 데이터 수집 및 이해(Data Acquisition and Understanding), 모델링(Modeling), 전개(Deployment), 인수(Customer acceptance) 등의 과정으로 구성된다.

〈표 1〉 데이터과학 역량을 다룬 선행연구 분석

항목	Conway(2010)	NIA(2014)	Zawadzki(2014) Modern Data Scientist	이명호(2016)
프로그래밍	• Hacking Skills	• 기술	• **Programming &** Database	• 프로그래밍
수학 및 통계; 분석	• Math & Statistics Knowledge	• 분석	• Math & Stat	• 통계
도메인 지식	• Substantive Expertise	• 사업	• **Domain Knowledge &** Soft Skills	-
커뮤니케이션 및 윤리	-	• 기반	• **Communication &** Visualization • Domain Knowledge & **Soft Skills**	• 비즈니스
데이터 수집, 처리 및 모델링	-	-	• Programming & **Database**	• 데이터 형식 • 데이터 기술
시각화	-	-	• Communication & **Visualization**	-
항목	PCMI(2016)	Tang & Sea-Lim(2016)	강지혜(2016a)	Syracuse iSchool Data Science Competency(2019)
프로그래밍	• Computational Thinking	-	-	• Technical Analysis
수학 및 통계; 분석	• Mathematical Foundations • Statistical Thinking	• Mathematics and Statistics	• 분석, 처리, 가공	-
도메인 지식	-	-	• 데이터 수명주기 관련 및 계획	-
커뮤니케이션 및 윤리	• Communication, Reproducibility and Ethics	• Communication Skills	• 활용	• Visualization and **Communication**
데이터수집, 처리 및 모델링	• Data Description and Curation • Data modeling	• Information Skills	• 시스템 구축 및 데이터 수집 운영	• Capture and Organization
시각화	-	• Visualization Skills	-	• **Visualization** and Communication

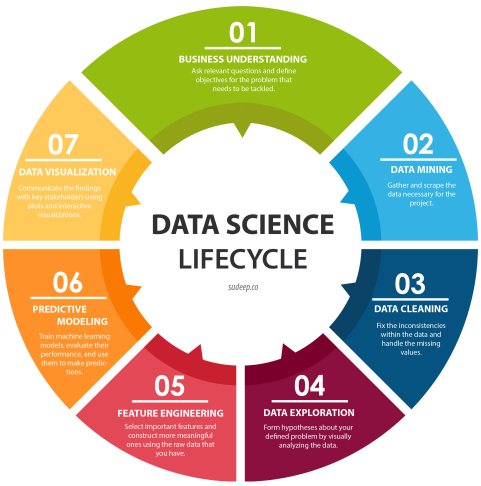
* 한 역량이 두 개의 요소로 조합된 경우, 관련 있는 것만 굵은 글씨로 표시함



〈그림 1〉 TDSP 데이터과학 생명주기 프로세스

Agarwal(2018)이 제시한 데이터터과학 생명주기는 다음과 같다(〈그림 2〉 참고). ① 프로젝트가 시작되기 전에 해결하려는 문제를 이해(Business Understanding)하고, ② 자료를 수집(Data Mining)하며, ③ 데이터를 정리하고 준비(데이터 정제, Data Cleaning)하는데 이 단계가 프로젝트 수행에서 가장 많은 시간을 필요로 한다. ④ 데이터 분석 전 브레인스토밍 과정인 데이터 탐색(Data Exploration)을 거치며, ⑤ 기계 학습에 필요한 데이터의 자질을 추출

하고 정리(Feature Engineering)한다. ⑥ 기계 학습이 실제 일어나는 예측 모델링(Predictive Modeling)이 이루어지며, ⑦ 커뮤니케이션, 심리학, 통계 및 예술 분야의 지식을 동원하여 데이터를 업무와 연결하고 문제를 해결하는 것으로 해당 프로젝트의 목표 달성을 증명하는 단계로 데이터 시각화(Data Visualization)가 진행된다.

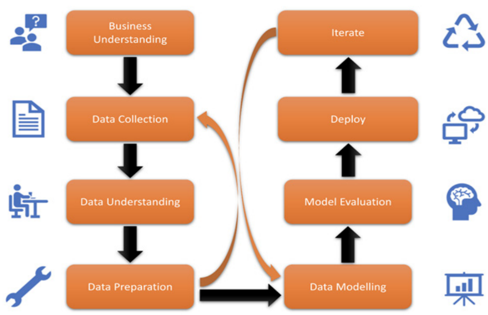


〈그림 2〉 Agarwal의 데이터과학 생명주기

〈그림 3〉은 Valiance Solutions의 데이터과학 프로젝트 생명주기 모형이다. 데이터과학 프로젝트 생명주기 모형을 통해 프로젝트 수행자는 문제를 식별하고, 활용 가능한 데이터 자원을 확인하며, 추가적으로 필요한 데이터 자원이 있는지 파악해야 한다. 정제된 데이터는 통계적 분석을 통해 문제 해결에 필요한 결과물로 구축되고 실행된 후 전개된다. 프로젝트의 결과는 시각화된 자료로 전달되고 성공한 프로그램은 고객에게 이전되고 그에 대한 교육이 실시된다. 마지막으로 지속적인 커뮤니케이션을 통해 유지보수가 진행된다.

〈그림 3〉 Valiance Solutions의
데이터과학 프로젝트 생명주기 모형

Sidhu(2019)의 모형은 〈그림 4〉와 같으며 문제를 인식하는 것 즉 업무를 이해하는 것으로 시작된다. 그 다음의 단계는 필요한 데이터를 수집하고 정제하여 모델링을 준비한다. 데이터 모델링은 정확률과 적합률을 기준으로 문제 해결에 필요한 솔루션을 제안한다. 구축된 모형이나 프로그램은 실제 업무에 적용하여 전개하고, 지속적인 피드백을 거쳐서 수정 및 보완이 이루어진다.



〈그림 4〉 Sidhu(2019)의
데이터과학 프로젝트 모형

<표 2>는 지금까지 살펴본 데이터과학 생명주기 모형들을 '문제인식', '데이터 수집 및 이해', '통계', '모델링', '실행 및 전개', '시각화', '커뮤니케이션', '유지보수'로 나누어 정리ㆍ분석하였다. 데이터 과학자의 역량 분석을 한 <표 1>과 함께 <표 2>도 4.2장의 데이터과학 역량을 기준으로 한 교과목 분석의 기초자료로 활용하였다.

3. 연구 방법

3.1 분석 대상

데이터과학 분야의 교육과정 분석을 위해 먼저, 대학의 공시 정보를 통합적으로 제공하고 있는 '대학 알리미(http://www.academyinfo.go.kr)'에서 2019년 현재 학과명에 '데이터'라는 단어를 포함하고 있는 학과를 개설ㆍ운영하고 있는 대학 및 대학원을 조사하였다. 1차 조사에서 얻어진 125개의 학교 및 학과 리스트 중에서 중복, 폐과 대상 및 데이터과학 분야와 연관성이 낮은 학과를 제외한 80개의 학과를 최종 분석 대상으로 선정하였다. 데이터과학 교육과정을 개설하고 있는 학교의 분류는 『교육통계연보』(2019)의 분류를 따랐다.

80개의 학과를 대상으로 교과목을 분석하기 위해 먼저 각 학교 및 학과의 홈페이지를 조사하였고, 홈페이지에서 정보를 얻기 어려운 학교에는 이메일, 전화 등을 활용하여 교과과정 정보를 수집하고자 하였다. 그러나 조사대상 중 2개 학과의 교과목을 수집할 수 없었는데, 이 중 1개 학교는 계약학과로 운영되어 교육을 의뢰한 기업의 비밀유지 요청에 의해 공개하지 않았다. 이 연구에서는 최종적으로 78개 학과의 총 2,041개의 교과목을 대상으로 분석을 실

<표 2> 데이터과학 생명주기를 다룬 선행연구 분석

항목	The Team DataScience Process(TDSP)	Sudeep Agarwal(2018)	Data Science Central	Life Cycle of a Data Science Project
문제인식	• Business Understanding	• Business Understanding	• Identify the Problem	• Business Understanding
데이터 수집, 처리, 준비, 이해	• Data Acquisition and Understanding	• Data Mining • Data Cleaning • Data Exploration	• Identify available Data Sources • Identify if additional Data Sources are needed	• Data Collection • Data Preparation
통계	-	• Feature Engineering	• Statistical Analysis	-
모델링	• Modeling	• Predictive Modeling	-	• Data Modelling
실행 및 전개	• Deployment	-	• Implementation & Development	• **Deploy** & Iterate
시각화	-	• Data Visualization	-	-
커뮤니케이션	• Customer Acceptance	-	• Communicate Results	-
유지보수	-	-	• Maintenance	• Deploy & **Iterate**

* 한 역량이 두 개의 요소로 조합된 경우, 관련 있는 것만 굵은 글씨로 표시함

시하였다. 교과목의 조사는 2019년 10월부터 두 달 간 진행하였다.

3.2 분석 방법

이 연구에서는 데이터과학 분야 교육과정 분석을 위해 다음과 같은 3가지 관점에서 분석을 실시하였다. 먼저, 학문 영역의 특징을 분석하기 위해 '대학알리미'에 공개된 정보를 이용하여 데이터과학 분야의 학과가 소속된 학교의 유형과 교육과정 운영 형식, 학문분류 현황을 조사, 분석하였다. 특히 학문분류 현황은, 각 학교에서 데이터과학 전공을 어떤 학문으로 분류하고 있는지 조사하기 위해 각 학과의 한국대학교육협의회(이하 대교협)의 교육편제단위 표준분류 정보를 조사하였다. 대교협의 표준분류 체계는 대학설립운영규정에 명시된 5대 계열(인문사회, 자연과학, 공학, 예체능, 의학)을 준용한 교육편제단위 분류체계로, 대분류-중분류-소분류 구성되어 있다(대학알리미 2019). 학문분류 현황의 분석을 통해 대학들이 데이터과학을 어떤 학문적 관점에서 접근하고 교육하고 있는지를 확인할 수 있다.

둘째, 데이터 전문가에게 요구되는 역량을 바탕으로 데이터과학 분야의 교과과정이 어떻게 개설・운영되고 있는지를 조사, 분석하였다. 이 연구에서는 데이터를 활용한 문제해결이라는 관점에서 데이터과학의 생애주기를 정의하고, 이를 바탕으로 데이터 전문가에게 요구되는 역량을 도출하였다. 그리고 이를 바탕으로 현재 운영되고 있는 교과과정을 매핑함으로써 역량을 기반으로 한 데이터과학 분야 교과과정의 특성을 분석하였다. 이러한 분석을 통해 데이터과학 교과과정이 데이터 전문가로서의 역량을 계발하는 데에 적합한 역할을 하는지 확인할 수 있다.

마지막으로, 교과과정의 주제적 특성을 확인하기 위해 교과목명에 나타난 키워드를 대상으로 빈도분석과 언어 네트워크 분석을 통한 내용 분석을 실시하였다. 언어 네트워크 분석을 위해서는 KrKwic(박한우, Leydesdorff 2004), UCINET 6.0(Borgatti, Everett and Freeman 2002)과 Netdraw(Borgatti 2002)를 활용하였다.

4. 데이터과학 교육과정 분석

4.1 학문 영역 특징 기반 분석

데이터과학 분야 교육과정을 운영하고 있는 학교의 유형별 개설 현황을 분석한 결과는 <표 3>과 같다. 데이터과학 전공은 전체 80개 개설 학교 중 4년제 대학(51.3%)과 대학원(47.5%)에 집중적으로 개설되어 있으며, 2019년 현재, 전체 4년제 대학(221개)의 약 18.6%, 전체 대학원(1,183개)의 3.2%에 해당하는 학교에서 데이터과학 교육과정을 운영하고 있는 것으로 나타났다.[1] 이것은 현재 우리나라에서의 데이터과학 전공 교육이 기술과 직업 실무적 관점보다는 학문적 접근을 바탕으로 한 연구 중심의 전문적 교육과정으로 자리 잡았음을 의미한

1) 현재 우리나라의 대학 및 대학원 과정의 수는 「교육통계연보」(2019)를 통해 확인하였음.

다. 특히 대학원 중 일반대학원의 수만큼 특수 대학원 과정이 개설되었다는 것은 4차 산업혁명을 중심으로 한 산업 구조의 변화에 따라 데이터 전문가 양성을 위한 계속교육에 대한 사회적 요구를 반영한 것으로 볼 수 있다.

〈표 3〉 학교 유형별 개설 운영 현황

학교 유형		개설 수	비율(%)
기능대학		1	1.3
소계		1	1.3
대학	일반대학	40	50.0
	사이버대학	1	1.3
소계		41	51.3
대학원	일반대학원	18	22.5
	전문대학원	4	5.0
	특수대학원	16	20.0
소계		38	47.5
합계		80	100.0

〈표 4〉는 학부과정을 대상으로 운영형식별 개설 현황을 조사한 결과를 나타내고 있다. 학부과정이 설치되어 있는 41개교 중 약 30%에 해당하는 12개교가 연계・융합전공의 형태로 데이터과학 교육과정을 운영하고 있는 것으로 나타났다.

〈표 4〉 학부(4년제) 운영형식에 따른 개설 운영 현황

학부 운영형식	개설 수	비율(%)
연계・융합전공	12	29.3
단일 전공	29	70.7
합계	41	100.0

연계・융합전공이란 둘 이상의 학과, 둘 이상의 학부 또는 학과와 학부 또는 둘 이상의 대학이 연계・융합하여 제공하는 전공을 말한다(「고등교육법시행령」(대통령령 제30131호) 제19조). 이 제도는 기존 학과를 그대로 둔 채 새로운 전공을 설치・운영할 수 있도록 만든 것으로, 대학의 탄력적 지식활동을 보장하는 학사운영의 자율적 제도 마련을 통해 융합・창의적 인재를 양성하는 것을 목표로 한다(교육부 2016). 학과 중심으로 운영되는 현재의 교육과정이 갖는 학문의 경직성 문제에서 벗어나 학과의 경계를 허무는 이러한 형태의 교육과정은 빠르게 변화하는 시대에 필요한 지식을 생성하고 확산하기 위한 제도 중 하나(서영진 2019)로, 연계・융합전공을 통해 대학은 교육 경쟁력을 제고할 수 있으며, 사회가 요구하는 인재상을 실현할 수 있다. 급속히 변화하는 4차 산업혁명 환경에서 사회가 요구하는 데이터 전문가를 양성하기 위해 대학에서는 학과 신설 외에도 다양한 학과 교과과정의 연계・융합을 통한 교육과정을 운영하고 있다. 특히 데이터과학은 데이터 처리에 필요한 통계, 수학, 기술적 지식 외에도 데이터 분석에 필요한 도메인 지식이 요구되는 학제적 성격을 갖고 있으므로, 데이터과학과 주제적으로 연관성이 있거나 데이터 분석 방법론의 적용이 요구되는 학과의 교과과정을 연계・융합함으로써 교육과정의 탄력적 운영이 가능하다.

데이터과학 교육과정을 연계・융합전공 형태로 운영하는 대학을 대상으로 어떠한 학과들이 참여하는지에 대해 분석하기 위해 참여 학과를 대교협 표준분류체계에 매핑한 결과는 〈표 5〉와 같다. 분석 결과, 인문・사회계열, 자연계열, 공학계열의 학과들이 데이터과학 교육과정 운영을 위해 참여하고 있고, 그 중에서 언론・방송・매체학, 사회학, 문헌정보학 등이 포

〈표 5〉 데이터과학 연계·융합전공 참여 학과의 계열 현황

계열(대분류)	계열(중분류)	계열(소분류)
인문·사회계열	언어·문학계열	독어·문학(2)*, 국어·문학, 중국어·문학, 언어학
	인문학계열	문화·민속·미술사학, 역사·고고학, 인문N.C.E.**
	사회과학계열	언론·방송·매체학(3), 사회학(3), 문헌정보학(2), 행정학(2), 국제학(2), 소비자·가정자원, 도시·지역·지리학
	경영·경제계열	경영학(8), 경제학(2), 경영정보학, 경영경제N.C.E.
자연계열	수학·물리·천문·지구계열	통계학(5), 수학(2), 물리학(2)
공학계열	전기·전자·컴퓨터계열	전산학·컴퓨터공학(4), 정보·통신공학(2), 전기·전자·컴퓨터
	건설계열	도시공학(2)
기타		예체능N.C.E., 자연과학N.C.E., 중등자연과학교육(컴퓨터), 생명과학, 산업공학

* 괄호 안의 숫자는 학교 수를 의미함
** Not Classified Elsewhere, 어디에도 분류되지 않음

함된 사회과학계열이 가장 다양하게 참여하고 있는 것을 알 수 있다. 또한 데이터과학의 연계·융합전공에 가장 많이 참여하는 학과는 경영학, 통계학, 전산학·컴퓨터공학 계열 학과인 것으로 나타났다. 이러한 분석을 통해 데이터과학은 다양한 학문 분야에서 관심을 갖고 접근하고 있는 학제적 성격의 학문 분야임을 알 수 있다.

교육과정 관점에서 데이터과학이 어떠한 분야의 학문으로 정의되고 있는지를 확인하기 위해 대교협의 표준분류체계를 이용한 학문분류 현황을 분석한 결과는 〈표 6〉과 같다. 대분류를 기준으로 했을 때 공학, 인문·사회, 자연과학 계열 순으로 분류된 학과 수가 많은 것으로 나타났으며, 학부과정의 경우에는 자연과학과 공학 계열로 분류되어 있는 경우가 가장 많았고, 대학원 과정은 공학과 인문·사회 계열로 분류되어 있는 경우가 가장 많은 것으로 확인되었다. 이것은 학부과정에서는 자연과학과 공학을 중심으로 한 관련 이론과 기술적 접근을 중심으로 교육과정을 운영하고 있는 반면, 대학원 과정의 경우 이론과 기술의 사회적 활용을 중심으로 운영하고 있다는 것을 의미한다.

〈표 6〉 학교 유형별 데이터과학 교육과정 분류 현황(대분류 기준)

대분류 \ 개설현황	전체		기능대학과정		학부(4년제)과정		대학원과정	
	개설 수	비율(%)	개설 수	비율(%)	개설 수	비율(%)	개설 수	비율(%)
공학	33	41.2	1	100.0	14	34.1	18	47.4
인문·사회	25	31.3	0	0.0	10	24.4	15	39.5
자연과학	22	27.5	0	0.0	17	41.5	5	13.2
합계	80	100.0	1	100.0	41	100.0	38	100.0

데이터과학의 학문분류 현황을 분류 계층별로 살펴보면 <표 7>에서 보는 바와 같이, '공학-전기 · 전자 · 컴퓨터-전산학 · 컴퓨터공학', '인문 · 사회-경영 · 경제-경영정보학', '자연과학-수학 · 물리 · 천문 · 지구-통계학' 계열에 가장 많은 학과가 분류되어 있는 것을 알 수 있다. 세부적으로 살펴보면, 학부과정에서 단일 전공으로 운영되는 학과의 경우에는 '자연과학-수학 · 물리 · 천문 · 지구-통계학' 계열에, 연계 · 융합전공은 '인문 · 사회-N.C.E.-N.C.E.'와 '자연과학-수학 · 물리 · 천문 · 지구-통계학' 계열에, 대학원과정에서는 '공학-전기 · 전자 · 컴퓨터-전산학 · 컴퓨터공학'에 가장 많은 학과가 분류되어 있는 것을 확인하였다. 이를 통해 데이터과학은 단일 학문으로서의 특성을 갖고 있지 않고, 정보기술, 경영학, 통계학 등 다양한 학문의 융합적 접근을 바탕으로 교육과정이 운영되고 있음을 알 수 있다.

<표 7> 데이터과학 교육과정의 계층별 분류 현황

대분류	중분류			소분류		
	분류명	개설 수	비율(%)	분류명	개설 수	비율(%)
공학	N.C.E	4	12.1	N.C.E.	4	11.8
	산업 · 안전	3	9.1	산업공학	3	8.8
	재료	1	3.0	반도체공학	1	2.9
	전기 · 전자 · 컴퓨터	25	75.8	N.C.E.	4	12.1
				응용소프트웨어공학	5	14.7
				전산학 · 컴퓨터공학	12	35.3
				정보 · 통신공학	4	11.8
	소계	33	100.0	소계	33	100.0
인문 · 사회	N.C.E	3	12.0	N.C.E.	3	12.0
	경영 · 경제	19	76.0	N.C.E.	2	8.0
				경영정보학	12	48.0
				경영학	5	20.0
	사회과학	3	12.0	N.C.E	1	4.0
				국제학	1	4.0
				언론 · 방송 · 매체학	1	4.0
	소계	25	100.0	소계	25	100.0
자연과학	N.C.E	2	9.1	N.C.E.	2	9.1
	농림 · 수산	1	4.5	축산학	1	4.5
	수학 · 물리 · 천문 · 지구	17	77.3	N.C.E.	2	9.1
				수학	1	4.5
				통계학	14	63.6
	화학 · 생명과학 · 환경	2	9.1	생명과학	2	9.1
	소계	22	100.0	소계	22	100.0

4.2 데이터 전문가 역량 기반 분석

2.2장에서는 데이터 전문가에게 필요한 역량 도출과 데이터과학 생명주기의 프로세스별 핵심 개념을 분석하였다(〈표 1, 2〉 참고). 본 장에서는 〈표 1, 2〉를 참고하여 데이터 전문가가 갖추어야 할 9개의 역량을 〈표 8〉과 같이 개발하였다.

〈표 8〉에서 제시한 역량은 데이터과학 교과목을 분석하는 지표로 활용하였다. 분석 전에 해당 역량에 할당될 수 있는 교과목의 키워드를 추출하여 분석의 효율성과 정확성을 높였다. '정보환경 이해'는 전반적인 정보환경을 이해하는 것으로 교과목에 개론, 특론, 연구, 논문, 연구방법론, 인문학적 소양, 데이터통신, 컴퓨터운영체계 등이 포함된 교과목을 할당하였다.

'데이터과학 실습'은 데이터과학 교과목을 재설계하는 연구(한국정보화진흥원 2014; Gil 2014; Lyon et al. 2015; 강지혜 2016a; 강지혜 2016b; PCMI 2016; Tang and Sae-Lim 2016; Thomas and Urban 2018 등)에서도 강조된 것으로 프로젝트, 실습, 캡스톤, 현장기반, 종합설계, 기획, 사례분석, 세미나 등이 포함된 교과목을 추가하였다.

'문제인식'은 전문지식을 기반으로 하여 문제를 인식하고 업무를 이해하는 것으로 전공 교과목, 경영분야(특히 마케팅), 사용자분석, 창업 아이디어 개발 등이 포함된 교과목을 주로 배정하였다.

'데이터 수집'은 필요한 데이터를 식별하고 수집하는 것으로 해당 기관의 내부 데이터 및 외부 데이터를 모두 고려한 것이다. 데이터 수집, 데이터 식별, 데이터 모델링 및 설계, 공공데이터의 이해, 정보조직, 비정형데이터, SQL, 웹크롤링 등과 관련된 교과목이 '데이터 수집' 역량으로 연결되었다.

'데이터 정제'는 데이터에 대한 일반적인 개념 뿐만 아니라 특정 도메인의 데이터가 가지는 속성을 이해하는 능력이 필요하다. 그러므로 통계적 분석을 위한 데이터 전처리 과정을 습득하는 교과목이 이에 해당되며, 데이터 추출, 색인, 데이터베이스화, 정보검색 등과 관련된 교과목도 포함되었다.

〈표 8〉 데이터 전문가 역량

역량	주요 내용
정보환경 이해	전반적인 정보환경 이해
데이터과학 실습	업무 기반 실습 및 사례 분석
문제인식	전문지식을 기반으로 한 문제인식 및 업무 이해
데이터 수집	필요한 데이터를 식별하고 수집
데이터 정제	통계적 분석을 위한 데이터 전처리
통계적 분석	기초수학 및 통계적 분석
기술적 구현	IT 기술을 바탕으로 한 플랫폼 구축 및 프로그래밍
결과 해석과 적용	결과 해석 및 실무에 적용·전개
유지관리	지속적인 피드백 제공에 따른 수정 및 보완 데이터 윤리 및 보안

'통계적 분석'은 기초수학 이론을 바탕으로 통계적인 분석이 실행되는 교과목에 초점을 맞추었다. 그러한 교과목으로는 수학, 통계적 분석, 데이터마이닝, 기계학습, 딥러닝, 인공지능, 소셜미디어 분석, R, SPSS 등이었다.

'기술적 구현'은 IT 기술을 바탕으로 한 플랫폼 구축 및 프로그래밍 능력으로 기술적 구현, 사물인터넷, 플랫폼 구축, 프로그래밍 언어 및 실습, Hadoop, 플랫폼 등의 교과목이 해당되었다.

'결과 해석과 적용'은 결과를 해석하고 결과물을 실무에 적용하고 전개하는 것으로 인문학적 소양을 통한 결과해석, 문제해결(모델개발)의 적용, 커뮤니케이션, 현장 적용, 시각화, UI, UX 등을 포함한 교과목이 이에 포함되었다.

'유지관리'는 피드백 제공에 따른 수정 및 보완, 그리고 데이터 윤리와 보안에 집중하여 개발된 프로그램이나 소프트웨어가 지속적으로 업무에 활용되고 변화하는 정보기술에 유연하게 대처할 수 있도록 하는 것이다. '유지관리'는 지속적인 프로그램 운영, 데이터 보안, 데이터 보호, 데이터 윤리, 저작권 등을 다루는 교과목과 연결되었다.

<표 8>을 기준으로 데이터과학 학부와 대학원 과정의 교과목을 분석하였다. 분석 항목은 9개의 역량에 미분류를 추가하였다. 미분류는 역량으로 간주할 수 없거나 교과과정에 포함된 졸업, 취업, 창업 관련 교육, 영어를 비롯한 외국어 강의 등이 해당되었다.

<표 9>는 학부와 대학원 교과목을 대상으로 데이터과학 전공자에게 필요한 역량을 분류한 결과이다. 먼저, 학부과정의 교과목에서는 36.4%로 통계적 분석 역량이 가장 높은 비율로 나타났다. 그 다음은 문제인식(15.6%), 정보환경 이해(10.3%), 기술적 구현(9.6%), 데이터과학 실습(8.1%) 순으로 나타났다. 가장 낮은 비율은 유지관리 역량으로 23개의 교과목(1.8%)만이 이에 해당되었다. 대학원과정 교과목의 분석 결과를 살펴보면, 가장 높은 비율을 차지한 것은 통

<표 9> 데이터 역량 기반 학부 및 대학원 교과목 분석

항목	학부과정		대학원과정	
	빈도	비율(%)	빈도	비율(%)
정보환경 이해	129	10.3	90	11.4
데이터과학 실습	101	8.1	46	5.8
문제인식	195	15.6	156	19.7
데이터 수집	37	3.0	27	3.4
데이터 정제	52	4.2	55	6.9
통계적 분석	455	36.4	271	34.2
기술적 구현	120	9.6	58	7.3
결과 해석과 적용	42	3.4	54	6.8
유지관리	23	1.8	24	3.0
미분류	95	7.6	11	1.4
합계	1,249	100.0	792	100.0

계적 분석 역량으로 34.2%이었으며, 그 다음으로는 문제인식(19.7%), 정보환경 이해(11.4%), 기술적 구현(7.3%), 데이터 정제(6.9%), 결과 해석과 적용(6.8%) 순으로 나타났다.

학부와 대학원 과정 모두 데이터 전문가의 역량으로 통계적 분석을 가장 중요하게 다루고 있음을 알 수 있었다. 또한 문제인식, 정보환경 이해, 기술적 구현 역량도 비중 있게 다루었다. 학부와 대학원과정을 비교해 보면, 학부과정에서는 기술적 구현, 데이터과학 실습, 통계적 분석 역량이 대학원과정에 비해 더 집중되었다. 이는 학부과정에서 수학 및 통계학적인 사고 함양, 프로그래밍 언어 습득, 현장 경험 등을 중요하게 다루고 있음을 보여준다. 반면, 대학원 과정에서는 문제인식, 결과 해석과 적용, 데이터 정제 역량이 학부과정보다 더 중요하게 다루어졌는데, 이는 대학원생이 학부생에 비해 해당 도메인의 전문지식, 데이터 자체에 대한 특성 파악 능력, 업무처리 경험 등을 두루 갖추고 있음을 보여주는 것이다.

4.3 교과목명 내용 분석

데이터과학 분야 교과과정이 어떠한 주제와 내용을 중심으로 구성되어 있는지를 확인하기 위해 수집한 교과목명 데이터를 이용하여 빈도 분석을 실시하였다. 단어 전처리 과정에서 동일한 개념의 이형표기는 하나로 통일하여 처리하였다. 〈표 10〉에서 보는 바와 같이, 전체적으로 가장 많이 출현한 단어는 '분석', '빅데이터', '통계', '데이터', '프로그래밍' 순으로, 이를 통해 '분석', '통계'와 '프로그래밍'이 데이터과학의 핵심 내용을 구성하고 있으며, 데이터과학의 연구 대상이 '빅데이터', '데이터', '정보' 등으로 혼재되어 표기되고 있음을 알 수 있다. 또한 '데이터마이닝', '데이터베이스', '기계학습', '시각화' 등이 데이터과학 분야의 주요 연구 분야이며, '경영', '수학', '컴퓨터'가 데이터과학 교과과정을 구성하는 주요 영역임을 확인할 수 있다.

교육과정별 특징을 살펴보면, 학부과정의 경우에는 '기초', '개론', '이해' 등과 같은 이론 중심의 교과목과 '실습', '응용', '캡스톤', '연습' 등 실습 중심의 교과목이 균형적으로 구성되어 있으며, 대학원과정의 경우에는 '특론', '개론', '기초' 등의 이론 과목과 '연구', '응용', '세미나' 등의 연구 과목으로 구성되어 있으며, '시스템', '네트워크', '인공지능', 'R', '딥러닝', '컴퓨팅' 등 기술적 접근을 통한 데이터과학 연구를 위한 교육과정을 제공하고 있는 것을 알 수 있다.

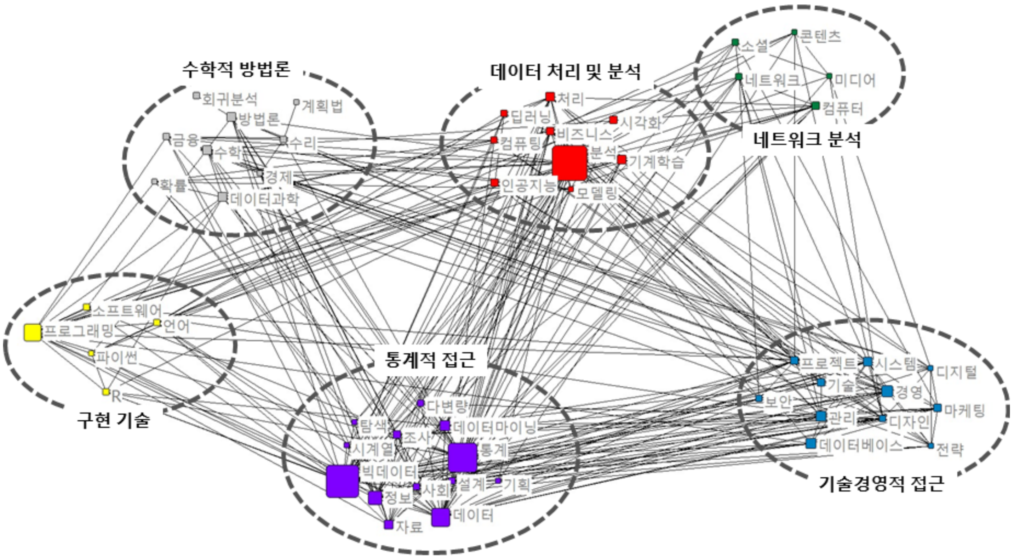
교과목명에 출현한 주요 키워드 간의 상관관계를 파악하기 위하여 UCINet의 CONCOR (CONvergence of iterated CORrelations) 분석을 실시하였다. CONCOR 분석은 네트워크 분석법 중 블록 모델링 방법 중 하나로, 구조적 등위성에 근거하여 노드들을 군집으로 나누는 방법이다(김용학, 김영진, 김영석 2008). 교과목명에 출현한 키워드에 대한 동시출현빈도 행렬을 이용하여 상관관계 분석을 수렴할 때까지 반복적으로 실행함으로써 연관성 있는 노드끼리 군집을 형성하고 이 군집 간의 관계를 나타냄으로써 복잡하게 얽힌 네트워크 안에 숨겨진 하위그룹간의 관계를 구조적으로 분석할 수 있다(박상훈, 이희정 2018).

이 연구에서는 데이터과학 교과목에 출현한 키워드 중 교과목의 주제와 거리가 먼, 교과목

〈표 10〉 데이터과학 분야 교과목명 키워드와 빈도(상위 30위)

<table>
<tr><th rowspan="2">순위</th><th colspan="2">전체</th><th colspan="2">학부과정</th><th colspan="2">대학원과정</th></tr>
<tr><th>단어</th><th>빈도</th><th>단어</th><th>빈도</th><th>단어</th><th>빈도</th></tr>
<tr><td>1</td><td>분석</td><td>309</td><td>분석</td><td>174</td><td>분석</td><td>133</td></tr>
<tr><td>2</td><td>빅데이터</td><td>282</td><td>통계</td><td>158</td><td>빅데이터</td><td>123</td></tr>
<tr><td>3</td><td>통계</td><td>245</td><td>빅데이터</td><td>155</td><td>통계</td><td>87</td></tr>
<tr><td>4</td><td>데이터</td><td>149</td><td>실습</td><td>94</td><td>데이터</td><td>65</td></tr>
<tr><td>5</td><td>프로그래밍</td><td>129</td><td>프로그래밍</td><td>90</td><td>특론</td><td>45</td></tr>
<tr><td>6</td><td>실습</td><td>103</td><td>데이터</td><td>83</td><td rowspan="2">연구, 정보</td><td rowspan="2">43</td></tr>
<tr><td>7</td><td>정보</td><td>102</td><td>정보</td><td>59</td></tr>
<tr><td>8</td><td>응용</td><td>82</td><td>기초</td><td>56</td><td rowspan="2">시스템, 프로그래밍</td><td rowspan="2">34</td></tr>
<tr><td>9</td><td>개론</td><td>78</td><td rowspan="2">개론, 응용</td><td rowspan="2">53</td></tr>
<tr><td>10</td><td>경영</td><td>75</td><td>데이터마이닝</td><td>33</td></tr>
<tr><td>11</td><td>기초</td><td>72</td><td>경영</td><td>50</td><td>응용</td><td>29</td></tr>
<tr><td>12</td><td>데이터마이닝</td><td>67</td><td>수학</td><td>41</td><td>처리</td><td>28</td></tr>
<tr><td>13</td><td>데이터베이스</td><td>61</td><td>자료</td><td>37</td><td>데이터베이스</td><td>27</td></tr>
<tr><td>14</td><td>관리</td><td>58</td><td>데이터마이닝</td><td>34</td><td rowspan="2">관리, 기계학습</td><td rowspan="2">26</td></tr>
<tr><td>15</td><td>연구</td><td>55</td><td rowspan="2">데이터베이스, 조사</td><td rowspan="2">33</td></tr>
<tr><td>16</td><td>시스템</td><td>54</td><td>경영</td><td>25</td></tr>
<tr><td>17</td><td>데이터과학</td><td>52</td><td rowspan="3">관리, 데이터과학, 캡스톤</td><td rowspan="3">31</td><td>개론</td><td>24</td></tr>
<tr><td>18</td><td>기계학습</td><td>50</td><td>비즈니스</td><td>23</td></tr>
<tr><td>19</td><td>자료</td><td>49</td><td rowspan="2">방법론, 시각화</td><td rowspan="2">22</td></tr>
<tr><td>20</td><td>처리</td><td>48</td><td>디자인</td><td>29</td></tr>
<tr><td>21</td><td rowspan="3">방법론, 수학, 특론</td><td rowspan="3">45</td><td>마케팅</td><td>28</td><td rowspan="2">데이터과학, 세미나</td><td rowspan="2">21</td></tr>
<tr><td>22</td><td>수리</td><td>27</td></tr>
<tr><td>23</td><td rowspan="2">금융, 연습</td><td rowspan="2">26</td><td rowspan="2">기술, 네트워크</td><td rowspan="2">20</td></tr>
<tr><td>24</td><td>시각화</td><td>42</td></tr>
<tr><td>25</td><td>조사</td><td>41</td><td>기계학습</td><td>24</td><td rowspan="2">R, 인공지능</td><td rowspan="2">19</td></tr>
<tr><td>26</td><td>마케팅</td><td>40</td><td rowspan="2">방법론, 설계</td><td rowspan="2">23</td></tr>
<tr><td>27</td><td>비즈니스</td><td>39</td><td rowspan="2">기초, 컴퓨터</td><td rowspan="2">15</td></tr>
<tr><td>28</td><td rowspan="2">세미나, 컴퓨터</td><td rowspan="2">37</td><td rowspan="3">언어, 영어, 이해, 컴퓨터</td><td rowspan="3">21</td></tr>
<tr><td>29</td><td rowspan="2">딥러닝, 컴퓨팅</td><td rowspan="2">14</td></tr>
<tr><td>30</td><td>프로젝트</td><td>34</td></tr>
</table>

의 특성을 나타내는 단어들(예: 개론, 실습, 연구 등)을 제외한 빈도 15 이상의 52개 상위 빈도어를 대상으로 CONCOR 분석을 실시하였으며, 〈그림 5〉와 같이 6개의 군집으로 구성된 네트워크를 형성하였다. 각각의 군집에 포함된 키워드의 주제적 유사성에 근거하여 군집을 대표하는 주제어를 부여한 결과, 데이터과학 교육과정은 크게 '데이터 처리 및 분석', '통계적 접근', '기술경영적 접근'과 같이 데이터과학의 학문영역을 구성하는 핵심 학문인 정보기술, 통계학, 경영학과 깊은 관련성이 있는 학문영역과 관련된 용어군과, 이들을 지원하는 '수학적 방법론', '구현 기술', '네트워크 분석'과 관련된 영역으로 용어군을 형성하고 있음을 알 수 있다.



〈그림 5〉 데이터과학 분야 교과목명의 CONCOR 분석 결과

5. 결론 및 제언

이 연구에서는 2019년 10월 현재 우리나라에 개설된 데이터과학 교육과정의 현황을 분석하기 위해 먼저, 데이터과학 분야의 교육과정을 분석한 기존 연구와 데이터과학 분야 전문가에게 요구되는 역량에 대한 분석을 진행하였고, 이를 바탕으로 우리나라에 개설된 80개의 교육과정을 대상으로 학문 영역 특징 기반 분석, 데이터 전문가 역량 기반 분석과 교과목명 내용 분석을 실시하였다. 분석 결과의 주요 내용은 다음과 같다.

첫째, 학문 영역의 특징을 기반으로 분석한 결과, 데이터과학 교육과정은 4년제 학부과정과 대학원에 집중적으로 개설된 것으로 나타났으며, 이는 현재 우리나라에서의 데이터과학 전공 교육이 기술과 직업 실무적 관점보다는 학문적 접근을 바탕으로 한 연구 중심의 전문적 교육과정으로 자리 잡았음을 의미한다. 또한 데이터과학의 학문분류 현황을 대교협의 표준분류체계를 기반으로 분석한 결과, '공학-전기 · 전자 · 컴퓨터-전산학 · 컴퓨터공학', '인문 · 사회-경영 · 경제-경영정보학', '자연과학-수학 · 물리 · 천문 · 지구-통계학' 계열에 가장 많은 학과가 분류되어 있는 것으로 나타났다. 이를 통해 데이터과학은 단일 학문이라기보다는 정보기술, 경영학, 통계학 등 다양한 학문의 융합적 접근을 바탕으로 다학제적인 교육과정으로 운영되고 있음을 확인하였다.

둘째, 데이터 전문가 역량 기반 교육과정의 특징을 분석하기 위해 먼저, 데이터과학 역량을 다룬 선행연구와 데이터과학 생명주기의 프로세스별 핵심 개념을 분석하여 데이터 전문가에게 요구되는 9개의 역량, 즉 '정보환경 이해', '데이터과학 실습', '문제인식', '데이터 수집', '데이터 정제', '통계적 분석', '기술적 구현', '결

과 해석과 적용', '유지관리'를 도출하였고, 이를 바탕으로 각 교육과정에서 개설하고 있는 교과목을 매핑하였다. 그 결과, 학부과정과 대학원과정 모두 데이터 과학자의 역량으로 '통계적 분석'을 가장 중요하게 간주하고 관련된 교과목을 가장 많이 개설하고 있는 것으로 나타났다. 또한 '정보환경 이해' 및 '기술적 구현' 역량도 비중 있게 다루고 있었다. 과정별 차이를 살펴보면, 학부과정에서는 '기술적 구현', '데이터과학 실습', '통계적 분석' 역량을 대학원보다 더 중요하게 다루고 있는 반면, 대학원과정에서는 '문제인식', '결과 해석과 적용', '데이터 정제' 역량을 학부에 비해 더 집중하여 다루고 있는 것을 알 수 있었다.

셋째, 데이터과학 분야 교육과정이 어떠한 주제와 내용을 중심으로 구성되어 있는지를 확인하기 위해 수집한 교과목명 데이터를 이용하여 빈도 분석을 실시하였다. 전체적으로 가장 많이 출현한 단어는 '분석', '빅데이터', '통계', '데이터', '프로그래밍' 순으로, 이를 통해 '분석', '통계'와 '프로그래밍'이 데이터과학의 핵심 내용을 구성하고 있음을 확인하였다. 또한 데이터과학 교과목에 출현한 키워드를 대상으로 CONCOR 분석을 실시한 결과, 데이터과학 교육과정은 데이터과학의 학문영역을 구성하는 핵심 학문인 정보기술, 통계학, 경영학과 깊은 관련성이 있는 학문영역과 관련된 용어군과, 이들을 지원하는 '수학적 방법론', '구현 기술', '네트워크 분석'과 관련된 영역으로 용어군을 형성하고 있음을 알 수 있었다.

'데이터'는 '정보', '지식'과 함께 문헌정보학 연구의 주된 대상임에도 불구하고, 데이터과학의 교육과정에 통계학, 컴퓨터공학, 경영학과 달리 문헌정보학적 접근이 두드러지게 나타나지 않은 것에 대해 생각해 볼 필요가 있다. 대규모 데이터로부터 정확한 통찰을 얻기 위해서는 데이터 품질에 대한 신뢰성이 확보되어야 하며, 그러므로 데이터 선정 및 평가를 아우르는 데이터 수집 및 데이터 정제의 과정은 매우 중요하다 할 수 있다(Liz et al. 2015). 데이터 전문가의 한 직종에 해당하는 데이터 사서(data librarian)는 데이터 자원에 대한 이해와 문제해결능력 및 통계적 지식을 바탕으로 데이터 분석을 지원하는 역할을 수행한다(Khan and Du 2018). 즉, 다양한 출처로부터 생산되는 데이터를 수집・정리하고 정제하여 데이터 분석으로부터 중요한 통찰을 얻을 수 있도록 지원하며, 이러한 업무 수행을 통해 데이터의 품질 관리 전반에 관여할 수 있다.

데이터 전문가에게 요구되는 역량 중, '데이터 수집'과 '데이터 정제'는 문헌정보학 교육과정에서 접근이 가능하다. 그럼에도 불구하고 현재의 교육과정에서는 이러한 역량을 기를 수 있는 교과목의 개설과 운영이 매우 부족한 실정이다. 그러므로 문헌정보학 교육의 관점에서 데이터 전문가를 양성하기 위해서는 데이터 기반 문제해결 능력에 기반하여 데이터 수집 및 정제 역량을 기를 수 있는 교육과정을 체계적으로 제공할 필요가 있다. 이는 기존의 문헌정보학 교육과정을 활용하거나 재설계함으로써 가능하다. 구체적으로는, 데이터의 식별과 수집, 자질 추출 및 정제, 전거제어 및 색인 이론 기반 데이터 표현, 데이터 공유 및 상호운용성 유지 및 설계, 데이터 보존 등의 역량을 기를 수 있는 교과목의 개발 및 설계가 필요하다.

본 논문은 2019년 10월을 기준으로 학부과

정 및 대학원과정에서 운영하는 80개의 데이터 과학 분야의 교육과정을 전수 조사한 연구라는 것에 의의가 있다. 이러한 현황 분석을 바탕으로, 현재 우리나라의 데이터 분석 실무 현장에서 요구되는 데이터 전문가의 역할을 확인하고, 한국형 데이터과학 교육 모형을 개발하는 후속 연구가 필요하다. 또한 이를 통해 데이터 전문가 양성을 위한 문헌정보학 교육의 역할을 규명할 수 있을 것이다.

참 고 문 헌

[1] 강지혜. 2016a. 문헌정보학과의 데이터 사이언스 커리큘럼 개발 실태와 방향성 고찰. 『한국도서관・정보학회지』, 47(3): 343-363.

[2] 강지혜. 2016b. 국가직무능력표준 직무분석에 따른 데이터 분석 교과 설계: 데이터 사이언스 교과 기초 능력단위요소에 대한 연구. 『한국문헌정보학회 하계 학술발표회』, 2016년 5월 27일. 부산: 부산대학교, 105-116.

[3] 교육부. 2016. 『창의혁신인재 양성을 위한 대학 학사제도 개선방안』. 세종: 교육부.

[4] 김용학, 김영진, 김영석. 2018. 한국 언론학 분야 지식 생산과 확산의 구조. 『한국언론학보』, 52(1): 117-140.

[5] 박상훈, 이희정. 2018. 소셜 빅데이터를 이용한 전통시장 활성화 요인 도출 연구. 『서울도시연구』, 19(3): 1-18.

[6] 서영진. 2019. 연계전공 참여 학생 실태조사 - 융합교육과정 운영을 위한 기초 연구. 『교양교육연구』, 13(3): 229-247.

[7] 유사라. 2018. 메타리터러시 관점에서의 문헌정보학 전공 커리큘럼 진단연구. 『한국문헌정보학회지』, 52(2): 191-220.

[8] 이명호. 2016. 데이터 사이언스 교과과정에 대한 연구. 『한국비블리아학회지』, 27(1): 263-290.

[9] 이상철. 2019. 빅데이터 전공 교과과정 개발. 『한국콘텐츠학회지』, 17(2): 29-33.

[10] 한국정보화진흥원. 2014. 『빅데이터 커리큘럼 참조 모델 Ver 1』. 대구: 한국정보화진흥원.

[11] Agarwal, S. 2018. *Understanding the Data Science Lifecycle*. [online] [cited 2019. 12. 21.] 〈http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/〉

[12] Bussaban, K. and Waraporn, P. 2015. "Preparing undergraduate students majoring in Computer Science and Mathematics with Data Science perspectives and awareness in the age of Big Data." *Social and Behavioral Sciences*, 197: 1443-1446.

[13] Dhar, V. 2013. "Data science and prediction." *Communications of the ACM*, 56(12): 64-73.

[14] Gil, Y. 2014. "Teaching Parallelism Without Programming: A Data Science Curriculum for

Non-CS Students." *IEEE*, 42-48.

[15] Hayashi, C. 1998. "What is Data Science? Fundamental Concepts and a Heuristic Example." Data Science, Classification, and Related Methods, 40-51. doi:10.1007/978-4-431-65950-1_3

[17] iSchool@Syracuse. *Data Science Competency*. [online] [cited 2019. 12. 13.] 〈https://ischoolonline.syr.edu/academics/applied-data-science/#curriculum〉

[18] Khan, H. R. and Du, Y. 2018. What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries, IFLA WLIC 2018. [online] [cited 2020. 1. 9.] 〈http://library.ifla.org/2255/1/139-khan-en.pdf〉

[19] Liz L., Mattern, E., Acker A. and Langmead, A. 2015. "Applying Translational Principles to Data Science Curriculum Development." *iPres 2015*. [online] [cited 2020. 1. 3.] 〈https://pdfs.semanticscholar.org/e4fd/3239f0edac113870e431aeba4e30fe8c6b8c.pdf?_ga=2.242823298.1275549363.1580577901-585377589.1561282565〉

[20] PCMI(The Park City Institute). 2016. *Curriculum Guidelines for Undergraduate Programs in Data Science*. [online] [cited 2019. 12. 23.] 〈https://www.stat.berkeley.edu/~nolan/Papers/Data.Science.Guidelines.16.9.25.pdf〉

[21] Sidhu, R. 2019. *Life Cycle of a Data Science Project: The major steps involved in tackling a real-world data science problem*. [online] [cited 2020. 1. 2.] 〈https://towardsdatascience.com/life-cycle-of-a-data-science-project-3962b9670e5b〉

[22] Tang, R. and Sae-Lim, W. 2016. "Data science programs in U.S. higher education: An exploratory content analysis of program description, curriculum structure, and course focus." *Education for Information*, 32(2): 269-290.

[23] The Team Data Science Process(TDSP). *The Team Data Science Process Lifecycle*. [online] [cited 2020. 1. 11.] 〈https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/lifecycle〉

[24] Thomas, C. V. L. and Urban, R. J. 2018 "What Do Data Librarians Think of the MLIS? Professionals' Perceptions of Knowledge Transfer, Trends, and Challenges" *College & Research Libraries*, 79(3): 401-423.

[25] Valiance Solutions. *Lifecycle of Data Science Projects*. [online] [cited 2019. 12. 29.] 〈https://valiancesolutions.com/wp-content/uploads/2018/03/Valiance-Insights_Mar_18_2.pdf?utm_source=Quora&utm_content=ebook+Lifecycle+DS+models〉

[26] Zawadzki, K. 2014. *Is Data Science a buzzword? Modern Data Scientist defined: Marketing Distillery*. [online] [cited 2020. 1. 2.]

<https://mywebvault.wordpress.com/2017/05/18/is-data-science-a-buzzword-modern-data-scientist-defined-marketing-distillery/>
[27] Zhang, J., Fu, A., Wang, H. and Yin, S. 2017. "The development of data science education in China from the LIS perspective." *International Journal of Librarianship*, 2(2): 3-17.

• **국문 참고자료의 영어 표기**
(English translation / romanization of references originally written in Korean)

[1] Kang, J. H. 2016a. "Study on the Current Status of Data Science Curriculum in Library and Information Science and its Direction" *Journal of Korean Library and Information Science Society*, 47(3): 343-363.
[2] Kang, J. H. 2016b. Study on the Competency of Data Science Curriculum. Korean Society for Library and Information Science Society Occasional Papers Series, May 2016, 년 5월 27일, Pusan: Pusan National University, 105-116.
[3] Ministry of Education. 2016. *Improvement of Undergraduate Degree for Raising Talents with Creativity-Innovation*. Sejong: Ministry of Education.
[4] Kim, Y. H, Kim, Y. J. and Kim, Y. S. 2008. "The Structure of Production and Diffusion of Knowledge in Korean Communication Studies." *Korean Journal of Journalism & Communication Studies*, 52(1): 117-140.
[5] Park, S. H. and Lee, H. C. 2018. "The Traditional Market Activation Factor Derivation Research through Social Big Data - Focused on Seoul City Mangwon market and Suyu market." *The Seoul Institute*, 19(3): 1-18.
[6] Seo, Y. Jin. 2019. "A Student Survey on Interdisciplinary Major - A Preliminary Study for the Implementation of Convergent Curriculum." *Korean Journal of General Education*, 13(3): 229-247.
[7] Yoo, S. R. 2018. "A Diagnostic Analysis of LIS Curriculum from the Meta-literacy Perspective." *Journal of The Korean Society For Library and Information Science*, 52(2): 191-220.
[8] Yi, M. H. 2016. "A Study on the Curriculums of Data Science." *Journal of The Korean Biblia Society For Library and Information Science*, 27(1): 263-290.
[9] Lee, S. C. 2019. "Development of Big Data Curriculum in University" *The Korea Contents Association Review*, 17(2): 29-33.
[10] National Information Society Agency. 2014. *NIA Bigdata Curriculum Reference Model*. Daegu: National Information Society Agency.