

## AI·데이터 사이언스 분야 직무 역량 강화를 위한 커리큘럼 연구\*

김효중\*\*, 김희웅\*\*\*

### 요약

4차 산업혁명에 따라 AI, 데이터 분석가 등 AI·데이터 사이언스 분야의 일자리에 대한 수요와 관심이 증가하고 있다. 그에 발맞춰 효과적으로 해당 분야의 직무를 수행할 수 있는 인력을 적시에 공급하기 위해서 구직자는 회사가 요구하는 역량을 개발하고, 대학은 양성 교육을 담당하여야 한다. 하지만, 적절한 역량을 갖춘 인력 공급의 이해 당사자인 구직자, 회사 그리고 대학 차원에서 적절한 대응 전략 마련에 어려움을 겪고 있다. 따라서, 본 연구는 필요충분한 직무 역량을 가진 인재 양성 및 공급을 위해 실무에서 요구되는 역량이 무엇인지 알아보고, 대학 차원에서의 역량 개발 방안을 제안하는 것을 목적으로 한다. AI·데이터 사이언스 분야에서의 필요 역량을 파악하고자 채용 플랫폼 링크드인(LinkedIn) 사이트의 채용공고 데이터를 텍스트 마이닝 기법을 활용하여 분석하였다. 이후, 국제적인 AI·데이터 사이언스 분야 대학원 교육과정과 채용 담당자와의 인터뷰 결과를 각각 토픽 모델의 결과와 비교 및 검증하는 절차를 통해, 대학 차원에서의 활용할 수 있는 커리큘럼을 제안하는 것으로 연구를 진행하였다.

주제어 : AI 직무 역량, 데이터 사이언스, 텍스트 마이닝, 네트워크 분석, 커리큘럼, 토픽 모델, 인터뷰

## A Curriculum Study to Strengthen AI and Data Science Job Competency\*

Kim, Hyo-Jung\*\* · Kim, Hee-Woong\*\*\*

### Abstract

According to the Fourth Industrial Revolution, demand for and interest in jobs in the field of AI and data science - such as artificial intelligence/data analysts - are increasing. In order to keep pace with this trend, and to supply human resources that can effectively perform such jobs in the relevant fields in a timely manner, job seekers must develop the competencies required by the companies, and universities must be in charge of training. However, it is difficult to devise appropriate response strategies at the level of job seekers, companies and universities, which are stakeholders in terms of supplying suitably competent personnel. Therefore, the purpose of this study is to determine which competencies are required in practice in order to cultivate and supply human talents equipped with the necessary job competencies, and to propose plans for the development of the required competencies at the university level. In order to identify the required competencies in the field of AI and data science, data on job postings on the LinkedIn site, the recruitment platform, were analyzed using text mining techniques. Then, research was conducted with the aim of devising and proposing concrete plans for competency development at the university level by comparing and verifying the results of the international graduate school curriculum in the field of AI and data science, and the interview results with the hiring managers, respectively, with the results of the topic model.

Keywords : AI job competency, text mining, network analysis, curriculum, topic model, interview

Received Feb 5, 2021; Revised Mar 25, 2021; Accepted Mar 30, 2021

\* This paper was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2018S1A3A2075114)

\*\* First Author, Master Course, Graduate School of Information, Yonsei University(libertygood9@yonsei.ac.kr)

\*\*\* Corresponding Author, Professor, Graduate School of Information, Yonsei University(kimhw@yonsei.ac.kr)

## I. 서론

### 1. 연구의 필요성 및 목적

AI 시대에는 일자리 구조에 근본적인 변화가 일어나며, 특히 단순한 업무의 자동화와 함께 창의성이 필요한 업무를 중심으로 새로운 일자리가 창출되는 등 직무변화 및 일자리 이동이 가속화되고 있다. 우선, 2020년 맥킨지 미래 전망 보고서에 따르면, 2030년까지 전 세계 일자리의 15~30%가 자동화되며, 한국의 경우 기존 일자리 700 만개가 감소하고 동시에 신규 최대 730 만개가 창출될 것으로 예상하였다(Mckenssey, 2020). 또한, 2020년 BCG 미래 전망 보고서에 따르면 비대면 디지털화가 빠르게 진행될 것이라는 전제 속에서 2030년까지 최소 10~50%의 일자리가 AI와 데이터 사이언스 관련 일자리로 대체되어 갈 것으로 밝혔다(BCG, 2020).

이처럼 급증하는 인력 수요에 발맞춰, 적시에 양질의 인재를 공급할 수 있는 환경 조성이 시급한 상황이다. 하지만, 직무 역량을 갖춘 인재 양성 및 공급의 이

해 당사자인 구직자, 회사, 그리고 대학 차원에서 서로 다른 관점을 가지고 있어 적절한 대응 전략 수립에 어려움을 겪고 있다. 우선, AI 및 데이터 사이언스 분야의 인력 수요에 발맞춰 해당 분야로 진출하고자 하는 구직자는 어떤 역량을 갖춰야 하는지 불투명한 상황이다. AI 관련 직군에 진출하고자 하는 수요는 급증하고 있지만, 구직자는 어떤 항목을 준비해야 할지 방향성을 잡지 못한다는 의견이 급증하고 있다고 밝히고 있다(KEIS, 2019).

한편, 회사 측 채용 담당자는 오히려 적절한 인력 확충에 어려움을 호소하고 있다. AI 및 데이터 사이언스 분야 중견기업 이상의 인사담당자의 53.5%가 신입사원 입사 후 업무 수행을 위해 재교육이 필요하다고 밝혔고, 이는 현재의 교육과정이 직무 수행에 충분하지 않다는 방증으로 볼 수 있다(KEIS, 2019).

직무 역량을 갖춘 양질의 인력 양성을 담당하고 있는 대학에서도 최적화된 교육 프로그램 개발을 위한 고민이 깊은 상황이다. AI 관련 직무에 종사하고자 하는 학생의 32% 이상이 대학의 정규 교육 프로그램이 아닌 외부의 사설 프로그램 등을 활용하여 역량을 준

〈표 1〉 AI·데이터 사이언스 분야 직무 역량 미스매치 분석  
(Table 1) AI·data Science job competency mismatch analysis

	2018		2019		2020	
	Job seekers scale(Ranking)	Recruiters scale(Ranking)	Job seekers scale(Ranking)	Recruiters scale(Ranking)	Job seekers scale(Ranking)	Recruiters scale(Ranking)
Internship	3.1(6)	4.2(1)	3.2(2)	4.1(1)	3.4(3)	4.2(1)
Leadership	3.2(4)	3.6(8)	3.1(5)	3.5(8)	3.1(8)	3.8(6)
Knowledge	3.3(2)	3.8(5)	3.4(1)	3.8(3)	3.5(2)	4.1(3)
Certificate	3.2(4)	3.7(6)	3.1(5)	3.5(8)	3.0(9)	3.7(8)
Prize	2.9(8)	3.7(6)	3.2(2)	3.8(3)	3.3(4)	4.0(4)
Project Experience	3.5(1)	3.9(3)	3.3(2)	4.0(2)	3.6(1)	4.2(1)
Problem Solving	3.0(7)	3.3(9)	3.1(5)	3.7(6)	3.2(6)	3.9(5)
Trendy	3.3(2)	3.9(4)	3.0(9)	3.7(6)	3.2(6)	3.6(9)
Communication	2.7(9)	4.0(2)	3.4(1)	3.8(3)	3.3(4)	3.8(7)

비하고 있다고 밝혔다(Choi & Shin, 2019).

AI 및 데이터 사이언스 분야에서의 상기 미스매치에 관한 이슈는 지난 3년간 크게 개선되지 않은 측면이 있다. <표 1>은 한국고용정보원에서 발간한 'IT서비스업 및 IT 기술 종사자의 직무 역량 미스매치 분석' 연구보고서에서 인용한 내용을 정리한 자료이다(KEIS, 2020). 이는 2018년부터 2020년을 포함하고 있으며, AI 및 데이터 사이언스 분야 연구 부분을 발췌·정리하였다. 다수의 구직자와 채용 담당자가 각 항목의 중요도를 5점 척도로 평가하고 이를 평균 낸 후 항목 간 순위를 선정하는 방식이다. <표 1>은 직무 역량 관련 중요도 인식 측면에서 구직자와 채용 담당자 간 미스매치가 발생하고 있다는 것을 확인하는 데 도움을 주고 있다.

한편, AI 및 데이터 사이언스 분야의 직무 역량에 관한 기존 연구 측면에서 다른 산업에 비해 활발하게 이루어지지 않은 측면이 있다. Deming and Kahn (2018)은 방대한 규모의 채용공고 데이터 분석을 통해 전체 산업군에 필요한 직무 역량을 도출하였지만, AI 및 데이터 사이언스 분야의 직무 역량 연구를 포함하지 않았으며, Durr, et al.(2020)은 AI 및 데이터 사이언스 분야의 직무 역량 관련 연구를 진행하였지만, 채용공고 및 커리큘럼 텍스트 분석에 그쳐 실제 활용성 측면에 아쉬움이 있다. 한편, Kim, et al.(2019)의 연구에서는 지원자의 이력서 기반으로 직무 역량을 도출하는 데 그쳐 실제 채용의 열쇠를 쥐고 있는 채용 담당자의 견해를 반영하지 못한다는 한계가 있다. 결론적으로 AI 및 데이터 사이언스 분야에 특화된 직무 역량에 대한 분석과 실제 현업에서의 관점, 그리고 이를 종합하여 대학에서의 커리큘럼을 보완하는 연구가 필요하다

따라서 본 연구는 AI 및 데이터 사이언스 분야에서의 업무를 원만하게 수행하기 위한 실무 역량을 파악하는 것을 목적으로 진행되었다. 추가로, 해당 업무에 종사하거나 채용하고자 하는 구직자와 회사, 그리고 인재 양성 프로그램을 개발하고 교육하는 대학이 가진 실무

역량에 대한 관점의 차이를 해소하는 데 이바지하는 것을 추구하고 있다. 본 연구를 통해 직무에서 필요로 하는 역량을 채용공고 데이터 기반으로 분석하여, 회사의 요구사항에 맞도록 구직자 그리고 대학에 유의미한 대응 전략을 제공할 수 있을 것으로 기대한다.

## 2. 연구의 방법 및 구성

본 연구는 크게 6단계로 구성된다. 첫째, 글로벌 채용 관련 플랫폼인 링크드인(LinkedIn)에서 'AI'와 '데이터 사이언스' 관련 영어 키워드를 활용하여, 채용공고 텍스트 데이터를 수집한다. 둘째, 수집된 채용공고 텍스트 데이터를 대상으로 토큰화, 불용어 제거, 워딩 복원, 명사 및 형용사 추출 등 전처리(Pre-processing)를 수행한다. 셋째, 전처리 과정을 거친 데이터를 바탕으로, 키워드 빈도분석, LDA(Latent Dirichlet Allocation) 토픽모델링, 의미론적 네트워크 분석, 그리고 연관어 분석을 진행하여 AI와 데이터 사이언스 분야에서의 실무에서 필요한 역량을 도출한다. 특히, 연관어 분석을 통해, 직급별 서로 다른 필요 역량을 추출한다. 넷째, 인재 양성을 담당하는 대학의 교육 프로그램을 확인하고 보완 사항을 제안하고자, US News Compass에서 선정한 AI-데이터 사이언스 분야 상위 5개 교육과정을 선정하여 교과목 데이터를 수집하고, 텍스트 분석결과 도출된 필요 역량과 비교 분석을 진행한다.

이후 채용공고 텍스트 분석결과로 도출된 직무에서 필요한 역량의 객관성을 확보하고자, 다국적 기업에서 AI와 데이터 사이언스 분야에 종사하며 채용에 관련한 경험이 있는 담당자 19인을 대상으로 인터뷰를 진행하여 비교 분석하고 검증하는 절차를 거친다. 마지막으로, 채용공고 텍스트 분석과 채용 담당자 인터뷰 결과를 기반으로, AI-데이터 사이언스 분야 5개 교육과정의 보완 요소를 제안한다. 마지막으로 본 연구의 의의와 한계점을 기술하고, 향후 연구과제를 제시한다.

## III. 선행 연구

### 1. 일반적인 산업에서의 직무 역량에 관한 연구

전 세계적으로 2008년 글로벌금융위기와 2020년의 코로나 19 팬데믹 상황을 겪으며 선진국을 중심으로 일자리 창출 문제 및 개인의 취업 문제가 사회적 화두로 자리 잡게 되었다. 이는 최근 5개년 간 채용공고 텍스트 데이터 혹은 인터뷰 방식으로 실무에서 필요한 직무 역량에 관한 연구가 쏟아지는 상황을 통해서도 확인할 수 있다. 대부분의 연구는 특정 산업에 최적화된 직무 연구가 아닌, 전체 산업군에 통용될 수 있는 포괄적인 직무 역량에 관한 연구가 많았다. Deming and Kahn(2018)은 십만 건에 달하는 채용공고 텍스트를 통해, 산업에서 요구되는 역량을 도출하였다. 그는 전체 산업군에서 통용되는 직무 역량을 10개의 핵심 토픽으로 제시하였다. Sibarani, et al.(2017)은 대용량의 채용공고 텍스트를 활용하여, 산업과 직급별로 필요 역량을 밝히는 데 이바지하였다. 한편, Kara, et al.(2017)은 다양한 산업군에 종사하고 있는 채용 담당자 인터뷰를 통해 직무 역량에 관한 연구를 진행하였다. 그 외에도 Kim, et al.(2019)은 취업 지원 시 활용한 자기소개서 데이터를 기반으로 회사의 채용공고에서 요구하는 역량과 비교 분석을 통해, 지원자와 채용 담당자의 시각적 차이를 확인할 수 있는 연구를 진행하였다.

상기 연구는 일반적인 산업에서 범용적으로 통용될 수 있는 직무 역량을 파악하는 데 연구의 의의가 있다는 점에서 공통점을 가지고 있다. 하지만, Deming and Kahn(2018)의 연구는 방대한 채용공고를 토픽별로 분류한 것에 그쳐, 채용의 이해관계자인 회사, 구직자 혹은 대학 관점을 포괄하지 못하였다는 한계가 있다. 추가로, Deming and Khan(2018)을 포함하여 일반적인 산업에서의 직무 역량에 관한 연구는 AI와 데이터 사이언스 분야의 직무 특수성을 반영하지 못한다는 측면에서 보완할 점을 가지고 있다. AI와 데이터

사이언스 분야는 기술과 비즈니스의 융·복합적 성격이 있으며, 지금까지 경험해 보지 못한 새로운 산업 패러다임을 구축해 나가야 한다는 측면에서 기존 산업과 상당한 차이가 있다는 것을 고려할 필요가 있다(Choi, et al., 2019).

### 2. 특정 산업에서의 직무 역량에 관한 연구

일반적인 산업에 통용되는 직무 역량에 관한 연구에서 최근에는 특정 산업의 특수성을 고려한 직무 역량에 관한 연구로 확장되고 있다. 특히, 특정 산업별로 서로 다른 채용공고 데이터 기반의 연구를 통해, 기존의 일반적인 산업에서 통용되는 직무 역량과 비교 분석을 진행하는 연구가 상당히 수행되었다. Lim, et al.(2020)은 작업치료사 직무 역량에 관한 연구를 관련 채용공고 데이터 기반 텍스트 분석을 통해 진행하였다. 해당 연구에서는 작업치료사 직무에서 필요한 역량이 무엇인지 파악하고, 추가로 일반적인 산업에서 통용되는 직무 역량 중 해당 직무에도 중요하게 작용하는 요소를 밝히는 방식으로 수행하였다. Kim, et al.(2018)은 소프트웨어 산업 분야에서의 채용공고 텍스트 데이터를 활용하여 요구되는 직무 역량을 확인하는 연구를 진행하였다. 해당 연구에서는 LDA 토픽모델링과 네트워크 분석을 통해 소프트웨어 직무에서의 필요 역량을 밝혔다.

상기 연구는 기존의 일반적인 산업 분야에 모두 통용될 수 있는 직무 역량을 밝히는 데에서 한 걸음 나아가 세부 직무별로 요구되는 역량을 밝히는 데 이바지하였다. 하지만, 세부 직무별로 필요한 역량을 제시하는 데 그쳐, 실제 해당 직무를 준비하는 구직자와 이들을 양성하는 대학 교육 프로그램의 보완 사항을 제안하는 데까지 나아가지 못하였다는 한계가 있다.

한편, AI와 데이터 사이언스 관련 직무에서의 필요 역량에 관한 연구도 최근 들어 진행되고 있다. Durr, et al.(2020)은 미국 Information 대학원의 교육과정과 AI 관련 채용공고 텍스트 데이터의 비교 분석을 진

행하였다. Schiler, et al.(2019)은 AI와 데이터 사이언스 분야에 취업을 희망하는 사람들의 이력서 텍스트 데이터를 수집하여 이를 통해 지원자 중심의 준비된 역량을 파악하는 연구를 수행하였다.

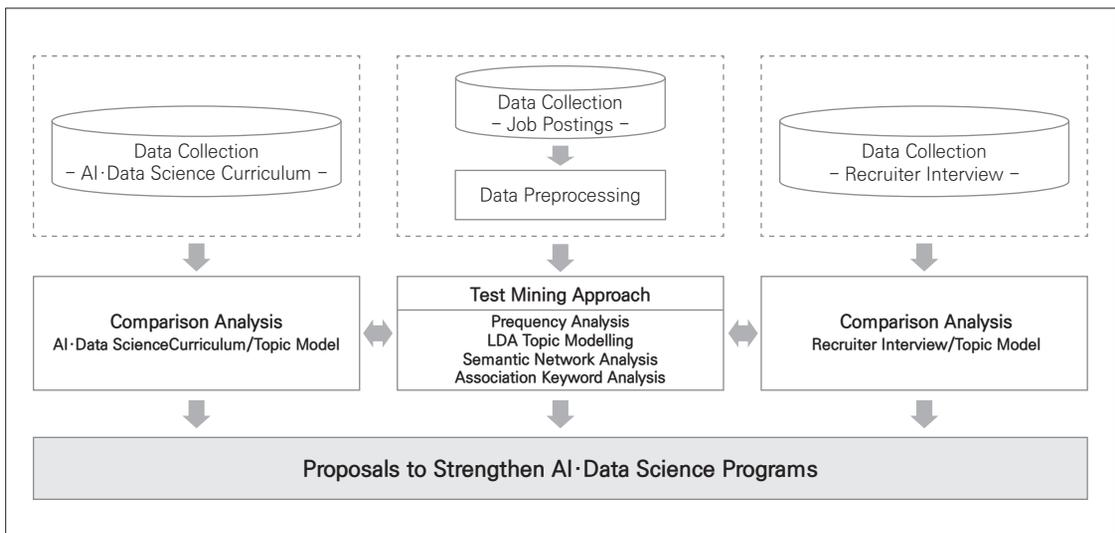
상기 연구는 AI와 데이터 사이언스 직무와 관련된 역량 연구가 수행되어 특정 산업의 특수성을 반영한 직무 역량을 파악하고 제시한 것에 의의가 있다. 하지만, Durr, et al.(2020)의 연구는 미국 Information 대학원의 교육과정과 AI 관련 채용공고 내 직무 역량과 비교 분석하는 데 그쳐, 교육과정의 보완 사항을 구체적으로 제안하는 것까지 나아가지 못한 한계가 있다. 또한, Schiler, et al.(2019)의 연구는 실제적인 채용의 열쇠를 쥐고 있는 회사 측에서의 요구 역량을 반영하지 못한 점이 한계로 남는다.

### III. 연구 설계

본 연구에서의 텍스트 분석은 글로벌 채용 플랫폼 링크드인(LinkedIn)에서 AI와 데이터 사이언스 영문 키워드로 데이터를 수집한 것부터 데이터 전처리, 키

워드 빈도분석, LDA 토픽모델링, 의미론적 네트워크 분석, 그리고 연관어 분석을 포함한다. 우선, 채용공고 내 키워드 빈도를 확인함에 전반적인 채용공고 텍스트의 경향을 파악하고자 한다. 다만 키워드 발생 빈도가 높다는 정보만으로는 채용공고의 구체적인 의미까지 연결짓기 어렵다. 따라서 LDA 토픽모델링을 활용하여, 문서 내 토픽 출현과 용어 출현을 확률적으로 모형화하여 핵심 역량을 파악할 필요가 있다. 추가로, 키워드 간 연결성을 확인할 수 있는 의미론적 네트워크 분석을 통해, 핵심 역량의 구체적인 의미를 해석하고자 한다. 끝으로, 서로 다른 직급을 모집하는 채용공고의 특성을 고려해 직급별 필요 역량의 차이를 구체적으로 제시하기 위해 연관어 분석을 진행하려고 한다.

텍스트 분석을 통해 도출된 직무 역량을 검증하고, 나아가 채용의 이해당사자인 대학의 역량 강화를 제안하는 것을 목적으로 채용 담당자 인터뷰를 진행한다. 인터뷰 질문에 대한 채용 담당자의 답변을 전사 기록한 후, 키워드 빈도분석을 통해 채용 담당자가 제공한 정보를 요약하여 제공하려고 한다. 채용 담당자 인터뷰는 금융, IT, 서비스, 제조업과 스타트업에서 AI와



〈그림 1〉 연구 프레임워크  
〈Fig. 1〉 Research framework

데이터 사이언스 관련 업무에 종사하며, 채용 경험을 가진 매니저급 이상으로 한정하여 수행한다.

텍스트 분석과 채용 담당자 인터뷰 분석을 거쳐 도출된 필요 역량을 기반으로 AI와 데이터 사이언스 전공 최상위 대학원 교육 프로그램의 교과목 구성을 확인한 후 보완점을 구체적으로 제시하는 것으로 마무리된다. 본 연구에 적용된 연구 프레임워크 도식화 자료는 <그림 1>과 같다.

## 1. 데이터 수집

본 연구의 주요 목적은 AI와 데이터 사이언스 분야 실무에서의 필요 역량을 관련 채용공고와 채용 담당자 인터뷰를 통해 구체적으로 제시하는 것이다. 추가로, AI·데이터 사이언스 분야 상위 교육 프로그램의 구성 교과목을 분석하여, 유의미한 인사이트를 도출하는 것을 목적으로 한다.

이러한 목적을 달성하기 위해, 본 연구에서는 채용공고 데이터, 채용 담당자 인터뷰 데이터, AI·데이터 사이언스 분야 상위 교육 프로그램의 교과목 구성 데이터를 수집하였다. 채용공고 데이터는 직무 역량 관련 텍스트 분석을 하기 위해 수집하였고, 채용 담당자 인터뷰 데이터는 분석 결과의 검증 및 신뢰성 확보를 위해 확보하였다. 교과목 구성 데이터는 실제 교육 현장과 실무와의 연계성을 확인하고자 수집하였다. 우선 전 세계 구인·구직 대표 플랫폼인 'LinkedIn'에서 2020년 4월 1일부터 5월 5일

까지 총 35일 간에 걸쳐 'AI'와 '데이터사이언스' 영문 키워드를 활용하여, 제목과 본문 내용을 포함한 7,350건의 채용공고 데이터를 수집하였다. 여기서 Staff와 Junior 직급을 모집하는 데이터는 3,895건(53%)을 차지하였으며, Manager 직급은 2,498건(34%), 그리고 Director급은 957건(14%)가 있었다(<표 2> 참조).

한편, US News Compass에서 선정한 AI·데이터 사이언스 분야 상위 5개 대학원 교육과정 프로그램을 선정하여, 각각의 교과목 데이터를 수집하였다. 또한, 다국적 기업에서 AI 및 데이터 사이언스 관련 업무에 종사하며, 채용 담당 업무를 경험한 19인의 인터뷰 데이터를 수집하였다. 인터뷰는 금융, IT, 서비스, 제조업 등의 산업군에 종사하는 이들로 구성하여, 산업간 필요 역량에 관해 비교 분석이 가능하도록 구성하였다.

## 2. 데이터 전처리(Pre-processing)

효과적인 텍스트 분석을 위해 전처리 작업을 진행하였다. 비정형 데이터를 분석하기 위해서는 전처리 과정이 필요하며, 전처리한 데이터는 분석의 신뢰성을 높일 수 있다.

<표 3>에서와 같이, Tokenization, Stopwords 제거, Lemmatization, 공통된 단어로 묶음, 추출 순으로 전처리를 진행하였다. 이 과정을 통해 비정형 데이터 특성에 따라 분석하기 어려운 형태로 존재하는 데

〈표 2〉 텍스트 데이터의 직급별 구성  
〈Table 2〉 Organization of text data by position

Position	Number of Postings	Ratio(%)
Staff / Junior	3,895	53
Manager	2,498	34
Director	957	14
Total	7,350	100

〈표 3〉 텍스트 데이터 전처리 단계 및 내용  
 〈Table 3〉 Text data preprocessing steps and contents

Preprocessing	Description
Tokenization	The process of separating words according to spaces
Stopwords	Remove unnecessary words using the Nltk toolkit
Lemmatization	Part of speech variants are readjusted to semantically identical words
Bundled into unified word	Unification of words that have the same meaning but expressed in different forms into one
Noun and adjective extraction	Extraction of meaningful parts of speech group for analysis

이더를 분석하기 쉽도록 체계를 구축할 수 있었다. 이후 정제된 데이터 중 명사와 형용사(분사 포함)만 추출하였다. 이는 영어 단어에서 사실상 의미를 전달하는 역할을 하는 품사 위주로 분석을 진행하기 위함이다.

한편, 채용 담당자 인터뷰는 국내 시가총액 상위 기업을 중심으로 다국적 기업인 회사에서 AI와 데이터 사이언스 업무에 종사하며, 채용을 담당했던 경험을 가진 매니저급 이상의 직책을 가진 이로 한정하여 진행하였다. 인터뷰는 대면 혹은 서면으로 진행하였으며, 대면으로 진행되는 경우 인터뷰 내용을 전사 기록한 후 연구에 활용하였다. 전사 기록된 인터뷰 데이터는 채용공고 텍스트 데이터와 마찬가지로 명사로 구성된 키워드로 가공하였으며, 영문으로 된 텍스트 분석 결과와의 비교 및 검증 절차를 진행을 위해 해당 키워드를 영문으로 변환하는 작업을 수행하였다.

### 3. 텍스트 마이닝 기법의 활용

본 연구에서는 채용공고 데이터를 텍스트 마이닝 기법을 활용하여 분석하고자 한다. 텍스트 마이닝 기법에는 LDA 토픽모델링, 감성 분석, 의미론적 네트워크 분석, 연관어 분석 등 다양한 방법이 포함되어 있다. 본 연구는 채용공고 데이터를 기반으로 실무에서의 핵심 역량을 도출하는 것을 연구 목적으로 두고 있으므로, LDA 토픽모델링 분석을 통한 토픽별 역량을 도출하는 것을 주요 기법으로 활용하고자 한다. 추가로, 의

미론적 네트워크 분석과 연관어 분석을 진행하며, 역량 간 의미를 깊이 이해하고 동시에 직급별 연관된 역량을 도출하여 세부적인 정보를 탐색하려고 한다.

LDA 토픽모델링은 문서나 텍스트(Corpus) 내에서 일정한 패턴을 찾아 잠재적으로 의미 있는 토픽을 발견하는 절차적 확률 분포 모델이다(Blei, 2012). 이는 어휘 모음이 특정 토픽으로 모인다고 가정하고, 해당 어휘가 토픽에 포함될 확률을 계산하여 그 확률이 높은 어휘 모음을 추출하는 방식으로 볼 수 있다(Blei, 2003). 즉, 문헌을 구성하는 단어들 독립적이지 않다는 가정하에 확률적으로 계산한 결과값을 각 토픽에 포함될 가능성이 큰 단어들의 집합으로 추출하는 알고리즘으로 정의할 수 있다.

LDA 토픽모델링에서는 어휘가 상호연관성이 있다는 전제 속에 어휘 생성 조건에 따라 사후확률을 추론한 후 LDA 그래프 모델을 통해 토픽을 도출하게 된다.

의미론적 네트워크 분석은 노드와 링크로 단어 간의 관계를 연결하여 구조를 수리적으로 분석하는 방법론이다(Newman, 2010). 이때 중심(Central) 노드를 기반으로 개별 노드의 역할, 위치, 영향력 등을 파악할 수 있다. Mimno, et al.(2011)은 LDA 알고리즘을 적용한 주제들 속에 공유하는 단어들을 중심으로 네트워크를 만들어 연결 중심성 및 근접 중심성이 높은 단어를 찾아내 연관성을 찾는 연구를 진행하였다. 본 연구에서 각각의 키워드의 연결선은 방향성을 가지지 않는 방식으로 분석을 진행하고, 이는 서로 영향을 미치

는 역량 키워드로 해석하려고 한다. 이때, 역량 키워드의 동시 출현 빈도가 높을수록, 큰 동시 출현 점수(Co-occurrence score)를 갖기 때문에, 해당 수치가 높은 키워드 동시 출현 묶음을 기반으로 연구를 수행하고자 한다. 의미론적 네트워크 분석을 통해, 역량 키워드 간의 관계를 통해, 직무 역량을 유기적으로 분석하고 의미 파악을 기대할 수 있다.

연관어 분석은 단어 사이의 관계를 파악해서 동시 출현 기반으로 분석하는 방법론이다(Blei, et al. 2003). 이때, 출현 단어 간 신뢰도 수치를 산출하여, 연관어 분석을 진행할 수 있다. 신뢰도는 항목 집합 X를 포함하는 출현 단어 중에서 항목 집합 Y도 포함하는 출현 단어 비율 말한다(Xianghua, et al., 2013). 본 연구에서는 직급의 차이가 가져오는 직무 역량의 차이를 파악하고자 해당 기법을 활용하려고 한다. 예를 들어, 채용공고 텍스트 데이터 문지에서 Manager와 신뢰도 수치가 높은 키워드를 도출하는 방식을 활용하여, Junior와 Staff, Manager, 그리고 Director 각각의 필요 역량을 파악하고 제시하고자 한다.

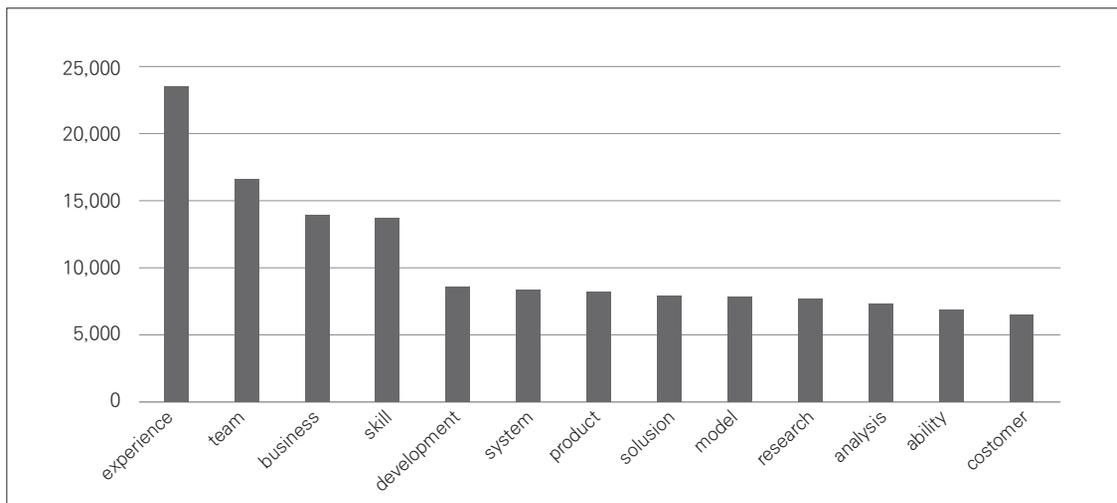
## IV. 데이터 분석

본 연구는 텍스트 마이닝 기법을 활용하여 채용공고 데이터를 분석하는 것을 시작으로 AI·데이터 사이언스 분야 교육과정 구성 교과목 데이터 분석과 채용 담당자 인터뷰 데이터 분석으로 나아가는 것으로 진행된다. 이때, 텍스트 마이닝 기법을 활용한 분석에는 키워드 빈도분석, LDA 토픽모델링, 의미론적 네트워크 분석, 그리고 연관어 분석 순으로 진행하여, 분석 단계를 거치면서 구체적으로 실무에서의 필요 역량을 파악하고, 직급별 요구하는 역량의 차이를 확인하는 것을 목적으로 분석이 수행된다.

### 1. 키워드 빈도 분석

링크드인(LinkedIn) 사이트에서 추출된 채용공고 데이터의 제목과 본문 내용을 이용하여 키워드 분석을 진행하였다. Python을 통해 전처리한 441,700개의 단어를 빈도순으로 정렬한 자료는 <그림 2>와 같다,

키워드 빈도를 살펴보면 experience가 압도적으로



<그림 2> 키워드 빈도 분석

<Fig. 2> Keyword frequency analysis

많이 등장함을 알 수 있다. 이를 통해, AI와 데이터 사이언스 직무에서는 해당 분야의 경험을 채용 과정에서 매우 중요하게 여기고 있음을 알 수 있다. 그와 함께 따라오는 team 역시 상기 분야의 특성상 팀 단위로 진행되는 프로젝트가 많아 협업 능력을 강조하는 측면이 있다고 해석해 볼 수 있다.

한편, business 키워드를 통해 AI와 데이터 사이언스 분야에서도 다른 산업군과 마찬가지로 비즈니스 마인드를 강조하고 있다는 사실을 알 수 있다. 또한, 상위군에 포함된 product, solution과 결합하여 AI 기술 그 자체를 넘어 실제로 비즈니스 문제를 해결하고, 수익을 창출할 수 있는 모델을 개발할 수 있는 역량이 필요하다고 볼 수 있다.

한편, skill, development, system, research, analysis, ability 등의 단어를 통해 볼 때, AI와 데이터 사이언스 분야의 주요 특성인 기술 중심성이 드러난다고 볼 수 있다. 기술을 개발하고, 시스템을 구축하며, 신기술을 연구할 수 있는 능력이 요구되는 것이다.

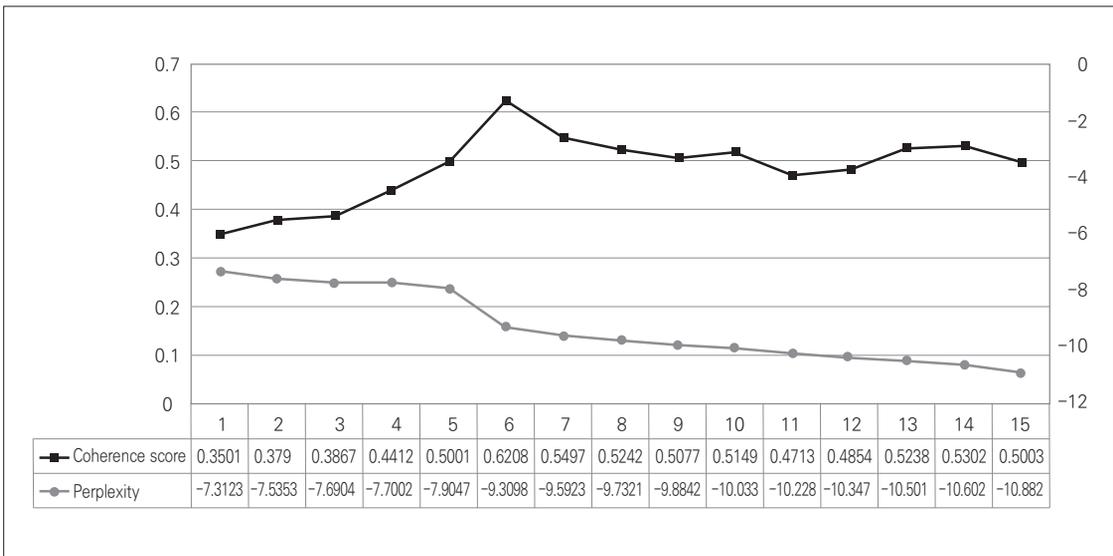
또한, customer의 등장 빈도가 상당히 후 순위에

밀려 있음에 주목할 필요가 있다. 일반적으로 제조업과 서비스업에서는 고객을 매우 중요시하는 측면이 있다. 특히, 영업 및 마케팅 등 고객과 더 밀접하게 대면하는 직군일수록 고객을 응대하는 능력 등이 강조되는 경향이 있는데, AI와 데이터 사이언스 산업에서는 상기 직군과 달리 비교적 고객과의 직접적 대면 측면에서 다소 떨어져 있음을 추론해 볼 수 있다.

## 2. LDA 토픽모델링

### 1) 토픽의 개수(K) 결정

LDA 토픽모델링 분석을 진행하기 위해서는 최적의 토픽 개수의 값을 결정해야 한다. 토픽의 개수에 따라 분석결과가 서로 다르게 나오기 때문이다. 따라서 최적의 토픽 개수를 정하는 데 도움을 주는 두 가지 평가지표를 활용하여, 최적 토픽의 개수(K)를 선정하고자 한다. Blei(2003)는 혼란도 지표를 활용하여, 낮은 혼란도를 제공하는 지점에서 토픽의 개수를 결정하는 것을 제안하였다. 하지만 최근 발표된 논문에서



〈그림 3〉 일관성 점수 및 혼란도 그래프  
 〈Fig. 3〉 Coherence Score and perplexity graph

낮은 혼란도가 데이터 분석가의 해석을 쉽게 하는 것은 아니라는 것에 대한 연구결과가 발표 되었다(Guo, 2017). 또한, 토픽의 개수가 증가함에 따라 혼란도는 낮아지기 때문에, 최적 토픽 개수(K)를 선정하는 데 어려움이 있다(Guo, 2017).

한편, 토픽 내 출현 용어들이 얼마나 해석하기 쉽게 구성되었는지를 평가하는 방법으로써, 해당 토픽에서 그 안에 구성된 단어가 유사한 맥락을 가졌는지 정도를 나타내는 일관성 점수를 통해 결정하는 방법을 활용할 수 있다(Newman, 2010). 즉, 이차적 데이터인 위키피디아를 활용하여 추출된 토픽의 질(Quality)과 비교하여 문맥적 일관성을 비교하여 결정하는 것이다(Lau, 2014).

본 연구는 혼란도와 일관성 점수를 모두 산출하여, 최적 토픽의 개수(K)를 산정하였다. 혼란도는 토픽의 개수가 증가함에 따라 낮아지는 경향이 있는데, 본 연

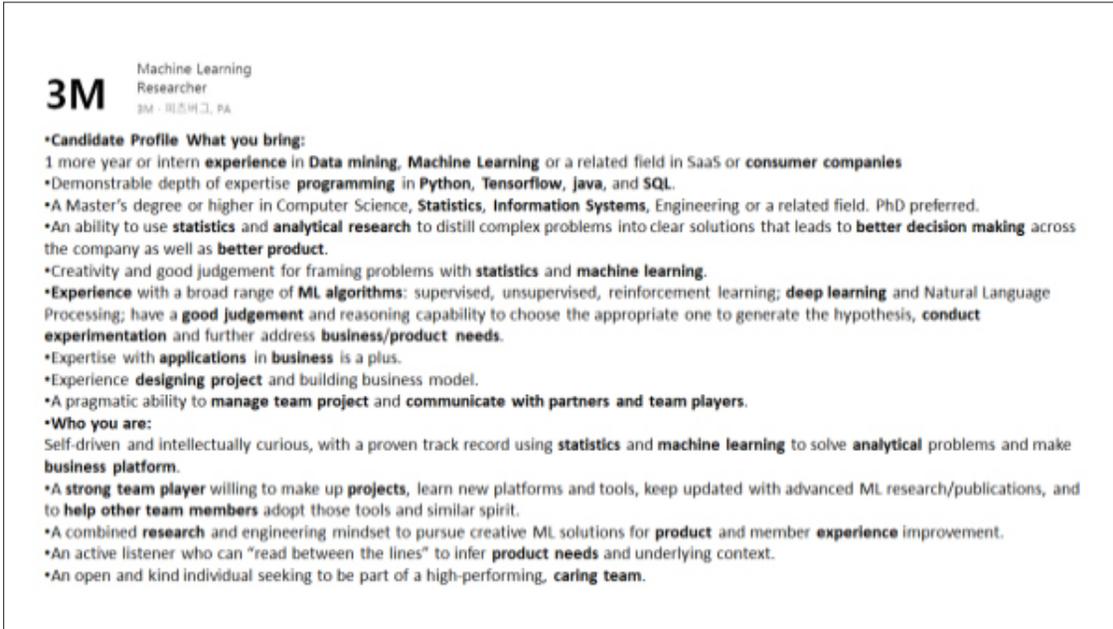
구에서 진행한 LDA 토픽모델링에서의 혼란도 지표에서는 특별한 변곡점 없이 통상적 추세선을 그대로 그리고 있어 특별한 시사점을 도출하기 어려웠다. 따라서, 혼란도의 하향 추세선과 일관성 점수 지표를 통해, 일관성 점수의 변곡점에서 최적 토픽의 개수(K)를 결정하였다(〈그림 3〉 참조). 최종적으로 6개의 토픽으로 결정하였으며, 일관성 점수는 0.6208로 산출되었다. 일관성 점수가 0.55가 넘는다면, 통상적으로 무난한 수준의 점수로서 분석을 진행하기에 무리가 없는 수준으로 평가할 수 있다(Newman, 2010).

## 2) 분석결과

LDA 토픽모델링 결과로 6개의 차원으로 데이터를 구분할 수 있었다. 각각의 토픽은 통계(Statistics), 프로젝트 경험(Project Experience), 머신러닝(Machine Learning), 비즈니스 문제해결(Business

〈표 4〉 LDA 토픽모델링 분석결과  
〈Table 4〉 LDA Topic modeling analysis results

	Topic	Keywords	Example sentence of the postings
1	Statistics	analytical, statistic, logical, research, decision, analysis, information	The ability to use statistics and analytical research to distill complex problems into clear solutions, which leads to better decision making across a given company as well as to better products.
2	Project Experience	experience, team, communication, considerate, interaction	Require 1 year or more intern experience in data mining, machine learning or a related team in SaaS or consumer companies.
3	Machine Learning	machine learning, deep learning, data mining, predictive modeling, visualization	Experience with a broad range of ML algorithms: supervised, unsupervised, reinforcement learning; deep learning and Natural Language Processing
4	Business Problem Solving	business, project, report, product, service, platform, problem, design	Expertise in using applications in business is a plus; experience of designing a project and building a business model.
5	Management	manager, leadership, people, control, applicant, partner, direction, status	Pragmatic ability to manage team projects and communicate with partners and team players.
6	Programming	Python, software, tool, java, technical, language, programming, solution	Demonstrable depth of expertise programming in Python, Tensorflow, Java, and SQL.



〈그림 4〉 채용공고 텍스트 데이터 사례 기반 키워드 분석  
(Fig. 4) Keyword analysis based on job posting text data case

Problem Solving), 프로젝트 관리 능력(Management), 프로그래밍 역량(Programming)으로 정의를 내릴 수 있다(〈표 4〉 참조). 한편, 토픽별 예시 문장은 개별 채용공고에서 일부 추출하여 제시하였다.

〈표 4〉에서 도출된 토픽은 크게 기술 중심 역량과 비즈니스 중심 역량 그리고 이를 아우르는 프로젝트 경험으로 정리할 수 있다. 기술 중심 역량으로는 통계, 머신러닝 그리고 프로그래밍 능력 토픽이 포함될 수 있고, 비즈니스 중심 역량에는 비즈니스 문제해결과 프로젝트 관리 능력이 포함된다고 볼 수 있다. 여기에 프로젝트 경험이 함께 강조된다는 것을 확인할 수 있다.

〈그림 4〉는 채용공고 데이터의 사례를 통해, 각 토픽에 연결 지을 수 있는 키워드를 구체적으로 확인할 수 있다. 또한, 실무에서의 관점으로 검증하기 위해 채용 담당자 인터뷰 분석을 수행하여 LDA 토픽모델링 분석결과의 적절성과 합리성을 확보하고자 한다.

### 3. 의미론적 네트워크 분석(Semantic Network Analysis)

LDA 토픽모델링을 통해 도출된 직무 역량과 관련된 토픽 및 키워드 간 연관성 정도를 확인해 보고자 네트워크 분석을 진행하였다(〈그림 5〉 참조). 네트워크 분석에 따른 노드(Node)의 크기는 문서 내에서 토픽에 이바지한 정도를 의미하며, 연결선을 가리키는 엣지(Edge)의 굵기는 연관성에 대한 강도를 의미한다(Kim, 2016).

상기 의미론적 네트워크 분석을 통해 연결성 정도를 파악할 수 있었다. 우선, 연결 중심성(Centrality)이 가장 강한 키워드는 experience임을 알 수 있다. 경험(experience)은 모든 토픽과 그 속에 소속된 키워드 간에 연결고리 역할을 하고 있다. 직무 역량에서 경험의 의미는 키워드 빈도분석, LDA 토픽모델링, 그리고 의미론적 네트워크 분석에 이르기까지 일관되게 가



〈그림 5〉 의미론적 네트워크 분석결과  
 〈Fig. 5〉 Results of semantic network analysis

장 중요한 의미가 있다는 데 주목할 만하다.

강한 연관성을 의미하는 굵은 선으로 연결된 관계를 중심으로 키워드 연관성을 파악하는 방식으로 진행하였다. 먼저, ‘experience-team-communication-project’로 이어지는 네트워크를 통해, AI와 데이터 사이언스 직무에서는 팀 단위의 프로젝트 경험과 상호 의사소통 역량이 중시된다는 것을 확인해 볼 수 있다. 한편, ‘experience-business’와 ‘experience-service’의 연관 관계를 통해, 프로젝트 경험을 통해 실질적으로 비즈니스와 서비스 개발로 연결할 수 있는 역량이 필요하다는 것을 도출할 수 있다. 즉, 프로젝트 경험 자체를 넘어 실질적으로 부가가치를 창출할 수 있는 능력을 보유한 사람이 실무현장에서 요구하는 인재라고 해석할 수 있다. 또한, 주목할 만한 키워드 연관 관계는 ‘experience-python’, ‘experience-software’ 그리고 ‘experience-technical’이다. AI와 데이터 사이언스 분야의 경험은 응용 소프트웨어를 활용하여 이루어지는 경우가 많으므로, 해당 기술을 자유자재로 활용하며 경험을 쌓아가는 역량이 필요하다고 볼 수 있다. 끝으로, ‘experience-analytical’과 ‘experience-research’의 관계를 통

해, 응용 소프트웨어 활용 능력과 더불어 현상을 분석하고 연구하여 비즈니스 문제를 해결할 수 있는 역량 또한 강조되고 있음을 알 수 있다.

요약하면, AI와 데이터 사이언스 분야에서 필요한 역량을 개발하기 위해서는 원만한 협업을 이룰 수 있는 사회적 기술, 데이터 분석 역량, 응용 소프트웨어를 활용할 수 있는 기술 그리고 비즈니스 모델로 구축해 나가는 능력을 종합적으로 포함하는 프로젝트 경험(experience)을 쌓는 것이 중요하다고 정리할 수 있다.

#### 4. 연관어 분석

본 연구에서 활용된 채용공고 데이터에는 다양한 직급에 대한 인력 수요를 담은 데이터가 포함되어 있다. 직급별 필요 역량에 대해 서로 다른 요소를 확인하고자 직급별 연관어 분석을 진행하는 것은 의미가 있다. 직급별 수행 업무와 역할이 다르므로, 각각 다른 역량 강화 방안을 구축할 수 있기 때문이다. 본 연구는 Apriori 알고리즘을 활용한 연관어 분석을 진행하였으며, 분석결과는 아래 〈표 5〉와 같다. 각각 신뢰도 점

〈표 5〉 연관어 분석결과  
 (Table 5) Results of related word analysis

Position	Counts (Ratio, %)	Related words
Staff / Junior	3,895 (53)	analytical, statistics, research, python, sql, ml, ***
Manager	2,498 (34)	teamwork, project, control, ml, experience, statistics, ***
Director	957 (13)	leadership, project, management, responsibility, revenue, business, ***

수 상위 기준으로 제시하였다.

우선, Staff와 Junior 수준에서는 통계적 분석 능력과 프로그래밍, 그리고 머신러닝 등이 강조된다고 볼 수 있다. 한편, Manager는 중간 관리자로서, 많은 역할을 담당하기를 요구받고 있다. Staff와 Junior 수준에서 필요하였던 통계, 머신러닝에 대한 기술적인 역량과 함께 팀과 프로젝트 관리 역량이 강조되고 있음을 볼 수 있다. 또한, 소통하고 통제하는 능력 또한 중요하다. 이는 실무 책임자로서, 팀을 이끌고 프로젝트를 관리할 수 있는 능력이 요구된다고 해석할 수 있다. 끝으로 Director 채용에서는 전체 조직에 대한 리더십을 발휘할 수 있고, 무엇보다 비즈니스 안목을 가지고 다양한 문제 상황을 해결할 수 있는 능력을 중심으로 평가하고 있다고 볼 수 있다.

## 5. AI·데이터 사이언스 커리큘럼과 토픽 모델의 비교 분석

AI과 데이터 사이언스 직무에 필요한 역량을 보유한 인력 양성을 담당하는 대학에서 이를 충분히 인지하고, 교육 프로그램에 충실히 반영하는지 살펴보는 것은 의미가 있다. 따라서, 본 연구에서는 LDA 토픽 모델 결과를 통해 확인할 수 있는 6가지 역량과 국제적인 AI·데이터 사이언스 커리큘럼과의 비교 분석을 통해, 유의미한 시사점을 도출하고자 분석을 수행하였다. 활용한 데이터는 US NEWS BUSINESS SCHOOL COMPASS에서 선정한 AI·Data Science Top School 중 상위 5개를 선정하여 교과목 데이터

를 수집하여 진행하였다. 비즈니스 역량에 중점을 둔 교육과정과 기술 역량에 중점을 둔 교육과정으로 양분되는 경향을 확인할 수 있었으며, 프로젝트를 경험할 수 있는 캡스톤 수업 혹은 산학협력 등의 교육과정은 다소 부족한 측면이 있었다.

〈표 6〉에서, 프로그램별 교과목을 정리한 후, 가장 많은 교과목이 포함된 토픽에 노란색, 다음은 연두색으로 표기하여 분석결과를 한눈에 알아보기 쉽게 정리하였다. 머신러닝과 관련된 교과목이 전체 커리큘럼 구성에서 상당한 비중을 차지하고 있는 것을 확인할 수 있다. 이는 전공 지식을 확립 차원에서 필수적인 요소라고 해석할 수 있다.

한편, Carnegie Mellon University와 University of Arizona의 경우에는 다른 커리큘럼에 비해 비교적 비즈니스 문제해결을 위한 교과목이 상당한 비중으로 구성되어 있음을 확인할 수 있다. 반면, University of Texas - Austin, University of Minnesota, 그리고 Georgia State University는 프로그래밍 역량과 머신러닝 등 기술 역량 함양에 더 중점을 두고 교육을 진행하고 있다고 평가할 수 있다.

한편, 일부 커리큘럼에는 통계 역량 강화를 위한 교과목이 전혀 개설되지 않기도 하였다. 또한, 모든 커리큘럼에서 프로젝트 경험을 지원하는 과목 개설에 소극적인 것을 확인할 수 있었다. 따라서, 현재 기술 중심 혹은 비즈니스 중심의 교육과정으로 구성된 커리큘럼을 상보적으로 개선하고, 프로젝트 경험을 할 수 있는 다양한 교과목 개설을 심도 있게 고민해 볼 필요가 있다.

〈표 6〉 AI·데이터 사이언스 커리큘럼과 토픽 모델의 비교 분석결과  
 (Table 6) Analysis of AI·Data science curriculum and topic model

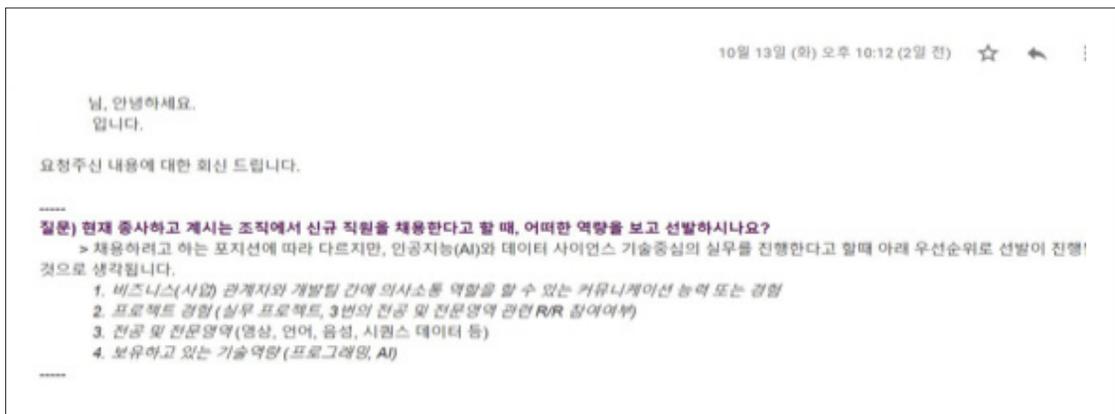
Topic	Carnegie Mellon University	University of Arizona		University of Texas – Austin		University of Minnesota – Twin Cities		Georgia State University			
		Business Consulting	10%	Business Foundations for IT Business Communication Business Data Networking Business Consulting Business Intelligence	35%	Capstones Applications for AI	16%	Business Essentials Optimization and Simulation for Decision	16%	Enterprise Architecture	8%
Project Experience	Data Analytics in Finance Capstone	10%	10%	Business Consulting	16%	-	-	-	-	-	-
Business Problem Solving	Business Fundamentals for Professionals Optimization for Prescriptive Analytics Business Value Through Integrative Analytics	37%	35%	Business Foundations for IT Business Communication Business Data Networking Business Consulting Business Intelligence	35%	Business Data Science	8%	Business Essentials Optimization and Simulation for Decision	16%	Enterprise Architecture	8%
Programming	Programming in R and Python	10%	20%	Software Design Big Data Technologies	20%	Advanced Programming & App Development Big Data & Distributed Cognitive Computing	20%	Programming for Data Management & Databases	16%	Introduction to Programming Application Programming Data Programming	24%
Statistics	Introduction to Probability and Statistical Foundations of Business Analytics	10%	-	-	-	-	-	Introduction to Statistics for Data Scientists	8%	-	-
Management	Managing Teams and Organizations Operations and Supply Chain Analytics	10%	10%	Enterprise Data Management	10%	Financial Management	8%	Building and Managing Teams	8%	Information Technology Project Management	16%
Machine Learning	Data Exploration and Visualization Machine Learning for Business Applications Modern Data Management	23%	25%	Web Mining & Analytics Information Systems Analysis & Design Data Mining for Business Intelligence	25%	Data Management User Generated Content Analytics Blockchain Technology Advanced Data Mining Financial Technology Advanced Blockchain	48%	Introduction to Business Analytics in R Exploratory Data Analytics and Visualization Predictive Analytics Big Data Analytics Ethics and Data Privacy	52%	Process Innovation Information Systems Security and Privacy Fundamentals of Database Management Unstructured Data Management Managing Big Data for IT Infrastructure for Big Data	52%

## 6. AI·데이터 사이언스 채용 담당자 인터뷰와 토픽 모델의 비교 분석

텍스트 마이닝 기법은 연구자의 주관적인 해석이 상당한 영향을 미치는 특징을 가지고 있다. 따라서, 본 연구에서 진행한 다양한 텍스트 분석결과를 실무에서 AI와 데이터 사이언스 업무를 수행하며, 실제 채용에 관여한 경험이 있는 채용 담당자와의 인터뷰 데이터와

비교하고 검증하는 절차를 통해, 텍스트 분석결과와 실제성과 정당성을 제고 하려고 한다.

채용 담당자와의 인터뷰에서 활용한 질문은 ‘AI·데이터 사이언스 분야 채용 시 중요한 고려사항이 무엇인가?’였으며, 담당자의 답변을 전사 기록한 후 키워드를 추출하여 각각 토픽 모델 결과와 연결지었다. <그림 6>에서 서면 인터뷰 진행 사례를 확인할 수 있다. 채용 담당자에게 답변을 받은 후 이를 명사 단위로



〈그림 6〉 서면 인터뷰 진행 사례  
 〈Fig. 6〉 Examples of written interviews

〈표 7〉 인터뷰 대상자의 소속 산업과 직급 정보  
 〈Table 7〉 Industry and position information of the interviewee

	Industry	Position
1	Finance	Senior Manager
2	Finance	Senior Manager
3	Finance	Manager
4	Finance	Manager
5	Finance	Manager
6	IT	General Manager
7	IT	Manager
8	IT	Manager
9	IT	Senior Researcher
10	IT	Research Engineer

	Industry	Position
11	Service	Senior Manager
12	Service	Senior Manager
13	Service	Manager
14	Service	Manager
15	Service	Manager
16	Manufacturing	Manager
17	Manufacturing	Manager
18	Startup	CEO
19	Startup	CTO
20	Startup	CEO

토큰화를 진행하고, 이를 각 토픽과 연결짓는 방식으로 분석을 수행하였다.

한편, 소속 산업별 서로 직무 역량을 파악하고자, 금융, IT, 서비스, 제조업 그리고 스타트업 등에서 19인을 섭외하여 진행하였다(〈표 7〉 참조). 〈표 8〉에서는

금융과 IT 분야에서 재직 중인 매니저 직급 이상의 채용 담당자와 인터뷰한 후 비교 분석한 것을 정리하였으며, 〈표 9〉에서는 서비스, 제조업, 그리고 스타트업 분야에서의 내용을 제시하고 있다.

〈표 8〉과 〈표 9〉에서는 채용 담당자별 답변 내용을

〈표 8〉 채용 담당자 인터뷰와 토픽 모델의 비교 분석결과(금융·IT)

〈Table 8〉 Comparative analysis of interview(Financial·IT) and topic model

	Project Experience	Business Problem Solving	Programming	Statistics	Management	Machine Learning	Others
K Bank	experience team experience	business service platform design	programming	statistics analytical	leadership	ML	degree
S Card	interaction intern team project experience	design	programming Python	statistics analytical research			degree
H. Life Insurance	experience team interaction	service platform design customer	Python	statistics	designing project managing		degree
M Securities	team player communication	application	programming		managing caring leadership	ML unstructured mining visualization	English
K. Life Insurance	team interaction intern team project	customer decision	Python	statistics		visualization predictive	
K. Mobility	team project team player communication interactive	customer decision	programming Python sql		managing		
S Telecom	experience team interaction intern	needs	programming Python	analytical research information		ML	written skills
L Telecom	interaction intern team project team player	judgement needs application	programming Python	statistics			
K Telecom	team player communication interactive	decision	java tool Software		managing	visualization predictive	certificates

〈표 9〉 채용 담당자 인터뷰와 토픽 모델의 비교 분석결과(서비스, 제조, 스타트업)

〈Table 9〉 Comparative analysis of interview(service, manufacturing, startup) and topic model

	Project Experience	Business Problem Solving	Programming	Statistics	Management	Machine Learning	Others
L. Dep	team project team player communication interactive	business service platform	programming	statistics analytical		predictive	certificates
H. H-shopping	team interaction intern	business service platform design	Python	statistics	leadership	predictive deep learning	MS office English
L. Mart	team project team player communication interactive	design customer decision	programming Python sql			ML	
E. mart	intern team project team player communication	judgement needs application	programming			visualization predictive	English
G. Retail	experience team interaction intern	customer decision judgement	Python		leadership control		
H. I&D	team project team player communication	service	programming	statistics analytical research information	leadership	visualization predictive	loyalty
POS.	team project team player communication	decision	programming Python sql				loyalty
Rocket *	experience team interaction intern	service platform	Python	statistics analytical research	leadership		
Rocket *	experience team interaction	business	programming Python Sql software	statistics analytical		ML	
Code *	experience team interaction	judgement needs application	Python	statistics	designing project managing		MS office

키워드로 정리한 후, 가장 많은 키워드를 포함한 토픽을 주황색, 다음은 하늘색, 세 번째를 보라색으로 표기하였으며, 토픽 모델의 결과 이외의 키워드는 별도로 표기하여 분석결과를 한눈에 알아보기 쉽게 정리하였다.

〈표 8〉에서, 금융과 IT 산업에서는 전반적으로 AI와 데이터 사이언스 관련 프로젝트 경험을 요구하고 있음을 알 수 있다. 하지만 금융 산업에서는 비즈니스 문제해결 역량과 통계와 프로그래밍 능력을 고루 중시하는 것에 비해, IT 산업에서는 프로젝트 경험을 대단히 중시하고, 프로그래밍 능력을 중요하게 평가하고 있다는 점에서 산업별 차이를 확인할 수 있다.

한편, 〈표 9〉에서와 같이 서비스업과 제조업, 그리고 스타트업도 서로 다른 관점을 가지고 있는 것을 확인할 수 있다. 서비스업은 프로젝트 경험과 비즈니스 문제해결을 중요하게 고려한다면, 제조업은 프로젝트 경험뿐 아니라, 프로그래밍과 통계 등 기술적인 역량을 강조하는 것이 특징이다. 스타트업은 프로젝트 경험, 기술적인 요소, 그리고 비즈니스 문제해결 역량을 두루 중시하는 경향을 보였다.

전반적으로, 머신러닝에 포함되는 각종 전공 지식에 대해서는 비교적 적게 언급되었다. 이는 프로젝트 경험 속에 자연스럽게 머신러닝 전공 지식이 활용되기 때문이라고 해석할 수 있다. 인터뷰 분석을 통해 본 연구에서 진행한 텍스트 분석결과를 검증하였다. 19인의 채용 담당자의 답변 내용을 키워드 중심으로 분석한 결과, LDA 토픽모델링을 통해 도출한 6가지 토픽에 전체적으로 포함된다는 것을 확인할 수 있었기 때문이다.

## V. 연구 토의 및 시사점

### 1. AI·데이터 사이언스 커리큘럼 방안 제안

본 연구에서는 키워드 빈도분석, LDA 토픽모델링, 네트워크 분석, 그리고 연관어 분석을 진행하였다. 추가로 다국적 기업의 채용 담당자의 인터뷰를 통해 검증한 결과를 바탕으로, 분석에 활용된 글로벌 상위 5

개의 AI·데이터 사이언스 커리큘럼(〈표 6〉 참조)에 관해 제언을 진행하고자 한다. 이후, 토픽 모델에서 도출한 토픽별 필요 역량을 기준으로 추가 개선을 고려할 만한 교과목 리스트를 구체적으로 제시하려고 한다. 이때, 교과목 리스트는 채용 담당자 인터뷰 및 글로벌 상위 5개 교육과정의 사례를 참조하여 구성하였다.

먼저, 분석에 활용된 AI·데이터 사이언스 커리큘럼 중에서 Carnegie Mellon Univ.와 Univ. of Arizona의 경우에는 각각 전체 교과목 중 절반 이상이 비즈니스 역량과 프로젝트 관리 부문에 쏠려 있는 경향이 있었다. 따라서, 통계, 프로그래밍, 그리고 머신러닝 등 기술 교육이 다소 부족한 측면이 있다. 이와 비슷한 교과목 구성을 가진 AI·데이터 사이언스 커리큘럼의 경우, 대용량 자료 처리 및 하둡, 스파크 관련 프로그래밍 개발 과목과 딥러닝, 강화학습 등 머신러닝 관련 역량을 보강할 필요성이 있다.

한편, Univ. of Texas-Austin, Univ of Minnesota-Twin Cities, 그리고 Georgia State Univ.의 경우에는 통계, 프로그래밍, 머신러닝 위주의 교과목이 구성되어 있었다. 하지만, 비즈니스 문제해결과 프로젝트 관리 등의 역량도 고루 개발할 수 있는 방향으로 고려해 볼 수 있다. 따라서, 해당 유형과 유사한 과목 구성을 가진 AI·데이터 사이언스 커리큘럼의 경우, 비즈니스 모델 구성과 조직 관리, 그리고 리더십 역량을 함양할 수 있는 교과목 개설을 고려해 볼 수 있다.

분석에 포함된 5개의 AI·데이터 사이언스 커리큘럼 모두 실무 경험을 할 수 있는 교과목이 다소 부족하였다. 이는 본 연구에서 진행한 다양한 분석에서 일관되게 강조되고 있는 프로젝트 경험 측면의 역량 개발을 위해 교육 프로그램 개설 측면에서 주의 깊게 관심을 가질 필요가 있다는 것을 시사한다.

지금까지 글로벌 상위 5개의 교육과정과 유사하게 구성된 다양한 AI·데이터 사이언스 커리큘럼에 대해 제언을 진행하였다. 지금부터는 실질적으로 개선을 고려할 수 있는 교과목을 채용 담당자 인터뷰 결과와 상위 5개 교육과정의 사례를 활용하여 제시하고자 한다.

〈표 10〉 AI·데이터 사이언스 커리큘럼 방안  
 〈Table 10〉 AI/data science curriculum plan

Topic	Proposed Courses	References
Project Experience	<ul style="list-style-type: none"> <li>• Internship with IT Company</li> <li>• Business Analytics Consulting</li> <li>• Corporate-linked research programs etc.</li> </ul>	<ul style="list-style-type: none"> <li>• IT industries' recruiter interview</li> <li>• Case of Univ. of Arizona</li> <li>• Overall recruiters' interview</li> </ul>
Solving Business Problems	<ul style="list-style-type: none"> <li>• Platform Business</li> <li>• CRM Data Analysis</li> <li>• Optimization &amp; Simulation etc.</li> </ul>	<ul style="list-style-type: none"> <li>• Finance industries' recruiter interview</li> <li>• Service industries' recruiter interview</li> <li>• Case of Univ. of Minnesota</li> </ul>
Management	<ul style="list-style-type: none"> <li>• Data Mining for Human Resources</li> <li>• Managing Teams and Organizations</li> <li>• Project Management etc.</li> </ul>	<ul style="list-style-type: none"> <li>• Overall industries' recruiter interview</li> <li>• Case of Carnegie Mellon Univ. and Univ. of Minnesota</li> <li>• Finance industries' recruiter interview</li> </ul>
Statistics	<ul style="list-style-type: none"> <li>• Regression(Graduate Level)</li> <li>• Mathematical Statistics</li> <li>• Multivariate Analysis etc.</li> </ul>	<ul style="list-style-type: none"> <li>• Overall industries' recruiter interview</li> <li>• Finance industries' recruiter interview</li> <li>• Manufacturing industries' recruiter interview</li> </ul>
Programming	<ul style="list-style-type: none"> <li>• SQL, Hadoop and Spark</li> <li>• Cognitive Computing</li> <li>• Data Structure and Algorithms etc.</li> </ul>	<ul style="list-style-type: none"> <li>• Overall industries' recruiter interview</li> <li>• Case of Univ. of Texas(Austin)</li> <li>• IT industries' recruiter interview</li> </ul>
Machine Learning	<ul style="list-style-type: none"> <li>• Reinforcement Learning</li> <li>• Advanced Blockchain</li> <li>• Web and App mining etc.</li> </ul>	<ul style="list-style-type: none"> <li>• IT industries' recruiter interview</li> <li>• Finance industries' recruiter interview</li> <li>• Service industries' recruiter interview</li> </ul>

〈표 10〉는 토픽별 필요 역량을 함양하기 위해 고려할 수 있는 교과목 목록과 이를 도출하기 위해 참고한 근거를 정리한 자료이다. 우선 첫 번째 역량인 Project Experience를 향상하기 위해 'IT Company-linked Internship', 'Business Analytics Consulting', 그리고 'Corporate-linked research program' 등의 과목 개설을 제안한다. 채용 담당자 인터뷰에서 특별히 강조한 인턴십과 산학 연계를 통한 실무 역량 증진 측면에서 필요한 과목으로 평가할 수 있다. 특히, IT 산업에 속한 채용 담당자는 IT 기업 인턴십이 필요하다고 역설한 점을 고려하였다. 또한, Univ. of Arizona의 컨설팅 회사와 연계한 인턴십 프로그램 운영 사례를 참조하였다.

두 번째 역량인 Business Problem Solving을 증진하기 위해 'Platform Business', 'CRM Data

Analysis', 그리고 'Optimization & Simulation' 등의 교과목 운영을 제안한다. 이는 금융과 서비스 산업 소속 채용 담당자의 인터뷰 과정에서 특별히 강조되는 요소를 반영하였으며, Univ. of Minnesota에서 제조 특성화 AI 인재 양성을 위해 최적화 이론 교과목을 운영하는 사례를 참조하였다. 세 번째 역량인 Management를 개발하기 위해 'Data Mining for Human Resources', 'Managing Teams and Organizations', 그리고 'Project Management' 등의 교과목 개설을 고려할 수 있다. 팀 운영과 프로젝트 관리 역량은 대부분 산업에 속한 채용 담당자 인터뷰에서 확인할 수 있었으며, 교과목은 Carnegie Mellon Univ.와 Univ. of Minnesota에서 해당 역량 함양을 목적으로 교육하고 있는 프로그램을 참조하였다.

한편, 또 다른 역량인 Statistics를 향상하기 위해 'Regression(Graduate Level)', 'Mathematical Statistics', 그리고 'Multivariate Analysis' 등의 교과목 개설을 제안한다. 이는 실무에서 특히 필요한 회귀분석, 수리통계학, 그리고 다변량 분석 등의 전공 지식을 요구하는 채용 담당자의 인터뷰에서 참조하였다. 다섯 번째 역량인 Programming을 증진하기 위해 'SQL, Hadoop and Spark', 'Cognitive Computing', 그리고 'Data Structure and Algorithm' 과목 운영을 제안한다. 대부분 산업에서 하둡과 스파크 등의 병렬 처리 프로세스에 대한 이해를 매우 강조하였으며, 그 외에도 컴퓨터 사이언스 기본 역량 함양을 목적으로 교육하는 교과목 사례를 참조하였다. 마지막 역량인 Machine Learning을 함양하기 위해 'Reinforcement Learning', 'Advanced Blockchain', 그리고 'Web and App mining' 과목 개설을 제안한다. 이는 IT, 금융, 서비스 산업에 속한 채용 담당자의 인터뷰에서 강조한 실무 역량을 참조하였다.

## 2. 시사점

### 1) 학술적 시사점

본 연구가 지닌 학술적 시사점은 다음과 같다. 첫째, 기존 채용공고 데이터 기반의 직무 역량 관련 연구(Kim, et al., 2019; Sibarani, et al., 2017)에서는 진행되지 않았던 채용 담당자 인터뷰와 AI-데이터 사이언스 커리큘럼 데이터 분석을 종합적으로 활용해 직무 역량을 구체적으로 제시하고 검증하는 방식으로 진행한 것이다. 이는 디지털 혼합방법론(Digital Mixed Methods) 연구로써, 텍스트 데이터 마이닝과 인터뷰의 혼합분석을 적용하여 분석의 합리성과 활용성을 확보하였다고 평가할 수 있다.

둘째, 기존 연구에서는 LDA 토픽모델링을 중심으로 채용공고 텍스트 분석에만 입각해 연구(Deming, et al., 2018; Kim, et al., 2019)를 진행하였다. 본 연구에서는 LDA 토픽모델링 이외에도 네트워크 분석,

연관어 분석 등 다양한 분석을 진행하였으며, 채용공고 데이터 이외에도 다양한 데이터를 복합적으로 활용하여 다면적 분석을 진행한 것에 의미가 있다.

셋째, 기존 연구에서는 산업 전반에 대해 포괄적인 직무 요건을 확인하는 연구는 진행되었지만, AI-데이터 사이언스 분야의 필요 직무 요건에 대해서는 제한적으로 연구가 진행되었다. 일부 진행된 연구 또한 채용공고와 미국 Information school 커리큘럼 데이터를 단순히 합쳐서 LDA 토픽모델링 분석(Durr, et al., 2020)을 진행하거나 채용 담당자의 관점이 반영되지 않은 지원자의 이력서만으로 직무 요건을 분석(Schiler, et al., 2019)하였다는 아쉬움이 있었다. 따라서 본 연구에서는 AI-데이터 사이언스 직무 역량을 중점적으로 연구하였다는 의의가 있다.

### 2) 실무적 시사점 및 향후 연구 방향

본 연구는 다음과 같은 실무적 시사점을 지니고 있다. 첫째, 요구되는 직무 역량을 탐색하고, 각 직무 역량에 따른 대응 전략을 대학 차원에서 구체적으로 제안한 점이다. 둘째, 대학 교육 측면에서 회사의 입장과 서로 다른 잘못된 의사결정으로 생길 비용을 줄이고 해당 분야의 노동 시장에 적절한 정보를 제공하였다는 것이다. 교육 프로그램을 설계하는 대학에서 실무에서 요구하는 역량을 양성하는 교과목 중심으로 운영할 수 있으며, 더불어 회사는 재교육 비용을 줄일 수 있다는 점에서 다자간 비용을 감소시키는 데 이바지하였다. 셋째, AI 및 데이터 사이언스 교육 커리큘럼을 비교 분석하면서, 각 프로그램의 특징과 강점을 파악하고 벤치마크로 활용하는 데 이바지한 것이다. University of Texas, Austin 커리큘럼은 AI 코어 기술 연구에 중점이 있지만, 최근 데이터 분석 기반 실제 비즈니스 가치를 창출하는 프로젝트 과목을 추가 개설하며 AI와 데이터 사이언스의 통합 역량 개발을 지향하고 있는 특징을 알 수 있다. 한편, 국내에서는 서울대학교 데이터 사이언스 대학원과 연세대학교 정보대학원 사례를 벤치마크 할 수 있다. 이들은 AI 코어 기술 연구뿐만

아니라 데이터 분석 기반 산업 연계성을 함께 연구하며 AI 및 데이터 사이언스 역량을 융·복합적으로 발전시켜 나가고 있다. 이처럼 국내외 대학원 프로그램 구성을 상세히 살펴봄으로써 서로의 강점을 벤치마크 하는 경험을 제공하였다는 것에 의의가 있다.

본 연구의 향후 연구 방향은 다음과 같다. 첫째, 본 연구는 AI·데이터 사이언스 분야의 직무 요건과 관련된 데이터를 활용하여 분석을 진행하였다. 하지만 AI 및 데이터 사이언스 분야는 공통적인 요소도 있지만 분리된 면도 있다. 따라서, 향후 각각 세분된 분야 및 다양한 직군을 대상으로 세밀한 분석을 진행한다면, 유용한 연구가 가능할 것으로 기대한다. 둘째, 본 연구는 2020년 상반기 중 한 달 정도의 기간에 걸쳐 텍스트 데이터를 수집하였다. 이는 연속된 기간의 데이터 변화를 분석하기 어렵다는 측면이 있다. 따라서 데이터 수집 기간을 연장하면 다채로운분석과 의의의 발굴이 가능할 것으로 기대한다.

## ■ References

- BCG (2020). "BCG on investing." March.
- Blei, D. (2003). "Latent dirichlet allocation." *Journal of machine learning research*, 3(4), 993-1022.
- Blei, D. (2012). "Probabilistic topic models." *Communications of the ACM*, 55(4), 24-27.
- Cho, S., Shin, S. & Kang, D. (2018). "A Study on the Research Trends on Open Innovation using Topic Modeling." *Informatization Policy*, 25(3), 52-74
- {조성배·신신애·강동석 (2018). 토픽 모델링을 이용한 개방형 혁신 연구동향 분석 및 정책 방향 모색. <정보화정책>, 25권 3호, 52-74.}
- Choi, B. (2016). "Study on Core Competency Improvement of University Students and Its Related Factors based on the Desired Talents of Companies." Doctor's Thesis, Sogyeong University.
- {최병삼 (2016). <기업의 인재상에 기초한 대학생의 핵심역량 향상과 관련요인에 대한 연구>. 서경대학교 박사 학위논문.}
- Choi, I. & Shin, E. (2019). "An Empirical Study of the Determinants of Successful Job Seeking of College Students." *Korea Economic Education Association*, 23(1), 23-49.
- {최일수·신은중 (2019). 대학 졸업자의 취업성과 결정 요인에 관한 실증연구. <한국경제교육학회>, 23권 1호, 23-49.}
- Deming, D. & Kahn, L. (2018). "Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for professionals." *Journal of Labor Economics*, 36(S1), 337-360.
- Desmet, P. & Hekkert, P. (2007). "Framework of Product Experience." *International Journal of Design*, 1(1), 13-23.
- Frey, C. & Osborne, M. (2017). "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting and Social Change*, 114, 254-280.
- Guo, Y., Barnes, S. & Jia, Q. (2017). "Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation." *Tourism Management*, 59, 467-483.
- Jung, Y. & Suh, Y. (2019). "Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews." *Decision Support Systems*, 123, 1-12.
- Karakatsanis, L., AlKhader, W., MacCroy, F. & Alibasic, A. (2017). "Data mining approach to monitoring the requirements of the job market: A case study." *Information Systems*, 65, 1-6.
- Kim, D., Kang, J. & Lim, J. (2016). "Comparative Analysis of Job Satisfaction Factors, Using LDA Topic Modeling by Industries : The Case Study of Job Planet Reviews." *Journal of Information Technology Services*, 15(3), 157-171.
- {김동욱·강주영·임재익 (2016). 토픽모델링 기법을 활용한 산업별 직무만족요인 비교 조사: 잡플래닛 리뷰를 중심으로. <Journal of Information Technology Services>, 15권 3호, 157-171.}

- Kim, J. & Huh, J. (2018). "Analysis of Occupation and Job Change in SW Field using Job Posting Data." *The Korean Institute of Information Scientists and Engineers*, 36(9), 15-21.
- {김정민·허정 (2018). 채용공고 빅데이터 기반 SW분야 직업 및 직무변화 분석. <정보과학회지>, 36권 9호, 15-21.}
- Kim, H. & Lee, H. (2016). "Trend Analysis of Data Mining Research Using Topic Network Analysis." *The Korean Society Of Computer And Information*, 21(5), 141-148.
- {김현희·이혜영 (2016). 토픽 네트워크 분석을 활용한 데이터 마이닝 분야 연구 논문 분석. <한국컴퓨터정보학회논문지>, 21권 5호, 141-148.}
- Kim, S. (2020). "A Study on the Similarity between the Self-introduction and the Talent of the company Using Text Mining." Master's Thesis, Soongsil University.
- {김세준 (2020). <텍스트 마이닝을 활용한 자기소개서와 기업의 인재상 유사성에 관한 연구>. 숭실대학교 석사 학위 논문.}
- Korea Employment Information Service(KEIS). (2019). *Future job prospects of the 4th industrial revolution*. Seoul: Korea Employment Information Service(KEIS).
- {한국고용정보원 (2019). <4차 산업혁명 미래 일자리 전망>. 서울: 한국고용정보원.}
- Korea Employment Information Service(KEIS). (2020). *Analysis of skill mismatch of IT technology workers in the IT service industry*. Seoul: Korea Employment Information Service(KEIS).
- {한국고용정보원 (2020). <IT 서비스업 IT 기술 종사자의 숙련 미스매칭 분석>. 서울: 한국고용정보원.}
- Lau, J. (2014). *Machine reading tea leaves: automatically evaluating topic coherence and topic model quality*. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics.
- Lee, J. (2020). "Forecasting Open Government Data Demand Using Keyword Network Analysis." *Informatization Policy*, 27(4), 24-46.
- {이재원 (2020). 키워드 네트워크 분석을 이용한 공공데이터 수요 예측. <정보화정책>, 27권 4호, 24-46.}
- Lee, S., Choi, S. & Kim, H. (2019). "An Exploratory Study of e-Learning Satisfaction: A Mixed Methods of Text Mining and Interview Approaches." *Information Systems Review*, 21(1), 39-59.
- {이순규·최수빈·김희웅 (2019). 이러닝 만족도 증진을 위한 탐색적 연구 : 텍스트 마이닝과 인터뷰 혼합방법론. <Information Systems Review>, 21권 1호, 39-59.}
- Lim, S., Yang, M., Han, D. & Park, J. (2020). "The Study of Job Market by Analysis of the Recruitment Advertisement for Occupational Therapy." *Therapeutic Science for Rehabilitation*, 9(2), 39-53.
- {임승주·양민아·한대성·박지혁 (2020). 채용공고 분석을 통한 작업치료사 구직 시장에 대한 연구. <재활치료과학>, 9(2), 39-53.}
- Lin, C. & He, Y. (2009). *Joint sentiment/topic model for sentiment analysis*. In Proceedings of the 18th ACM conference on Information and knowledge management.
- Mckensy (2020). "The recovery will be digital." August.
- Mimno, D. (2011). *Optimizing semantic coherence in topic models*. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing.
- Newman, M. (2010). *Networks : An Introduction*. Oxford University Press.
- Shin, J., Choi, J. & Koh, W. (2015). "A study on the Use of Learning Analytics in Higher Education: Focusing on the perspective of professors." *Journal of Educational Technology*, 31(2), 223-252.
- Steyvers, M. & Griffiths, T. (2007). "Probabilistic topic models." *Handbook of latent semantic analysis*, 427(7), 424-440.
- Xianghua, F., Guo, L., Yanyan, G. & Zhiqiang, W. (2013). "Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon." *Knowledge-Based Systems*, 37, 186-195.

Yun, Y. & Ji, H. (2015). "A development of Open Social Learning Platform for learning analytics and educational data mining." *Journal of Institute of Information Scientists and Engineers*, 23(12), 1349-1351.