

온라인 채용정보를 이용한 데이터 과학자 요구 역량 탐색

Exploring the Job Competencies of Data Scientists Using Online Job Posting

김향단(Xiangdan Jin)*, 백승익(Seung Ik Baek)**

초 록

4차 산업혁명으로 글로벌 비즈니스 환경이 빠르게 변화함에 따라 기존에는 없던 새로운 직종들이 등장하고 있다. 새롭게 등장한 직종 중에 최근에 기업들이 가장 많은 관심을 가지고 있는 직종은 '데이터 과학자 (Data Scientist)'일 것이다. 인터넷과 같은 정보통신 기술이 우리들의 생활에서 차지하는 비중이 커지면서 온라인에서의 활동 뿐만 아니라 오프라인 상에서의 활동에 대한 데이터가 매시간 컴퓨터에 저장되어 빅데이터를 생성하고 있다. 기업들은 이런 빅데이터로부터 새로운 기회를 창출하기 위하여 많은 노력을 기울이고 있다. 이런 기업의 노력과 함께 새롭게 등장한 직종이 바로 '데이터 과학자'이다. 빅데이터 시대를 이끌어갈 유망 직업인 데이터 과학자에 대한 수요는 끊임없이 증가되고 있지만 공급은 여전히 부족한 현황이다. 분석과 관련된 기술과 도구들이 새롭게 개발되고 있음에도 불구하고 기업은 여전히 이러한 기술을 목적에 맞게 활용할 수 있는 전문가를 찾는데 있어서 많은 어려움을 겪고 있다. 데이터 과학자 부족 문제를 심각하게 만드는 주요 이유 중 하나는 데이터 과학자 직무에 대한 이해가 부족하다는 점에서 찾을 수 있을 것이다. 이에 본 연구에서는 기업에서 필요로 하는 데이터 과학자의 역량을 기업의 실제 채용정보를 정성적으로 분석하여 보았다. 연구 결과, 과거 소프트웨어 엔지니어나 시스템 분석가들에게 요구되었던 Technical Skill과 System Skill 뿐만 아니라 비즈니스 컨설턴트나 Project Manager에게 요구되었던 비즈니스 관련 스킬이나 효율적인 팀워크를 위한 대인관련 스킬도 광범위하게 요구됨을 발견하였다. 본 연구결과를 통하여 데이터 과학자란 직업에 관심을 가지고 있는 사람들과 데이터 과학자를 채용하기를 원하는 기업에게 가이드라인을 제공하여 줄 것으로 기대하고 있다.

ABSTRACT

As the global business environment is rapidly changing due to the 4th industrial revolution, new jobs that did not exist before are emerging. Among them, the job that companies are most interested in is 'Data Scientist'. As information and communication technologies take up most of our lives, data on not only online activities but also offline activities are stored in computers every hour to generate big data. Companies put a lot of effort into discovering new opportunities from such big data. The new job that emerged along with the efforts of these companies is data scientist. The demand for data scientist,

* First Author, Graduate Student, School of Business, Hanyang University(hyangdan0129@gmail.com)

** Corresponding Author, Professor, School of Business, Hanyang University(sbaek@hanyang.ac.kr)

Received: 2022-01-27, Review completed: 2022-03-07, Accepted: 2022-03-23

a promising job that leads the big data era, is constantly increasing, but its supply is not still enough. Although data analysis technologies and tools that anyone can easily use are introduced, companies still have great difficulty in finding proper experts. One of the main reasons that makes the data scientist's shortage problem serious is the lack of understanding of the data scientist's job. Therefore, in this study, we explore the job competencies of a data scientist by qualitatively analyzing the actual job posting information of the company. This study finds that data scientists need not only the technical and system skills required of software engineers and system analysts in the past, but also business-related and interpersonal skills required of business consultants and project managers. The results of this study are expected to provide basic guidelines to people who are interested in the data scientist profession and to companies that want to hire data scientists.

키워드 : 데이터 과학자, 온라인 채용 정보, 직무분석, 직무역량 분석, 콘텐츠 분석, NVivo, Word Tree
Data Scientist, Online Job Posting, Job Analysis, Job Competency Analysis, Content Analysis, NVivo, Word Tree

1. 서 론

예로부터 기술의 발전은 기존 직업의 소멸과 새로운 직업의 탄생을 야기시켰다. 포드 자동차의 대량 생산으로 마부는 사라졌고, 운전기사와 자동차 정비공이 등장하였다. 컴퓨터와 인터넷의 등장은 소프트웨어 개발자, 프로그래머 등의 새로운 직업들을 탄생시켰고, 4차 산업 혁명은 이러한 직업의 변화를 더욱 가속화시켰다. 특히 빅데이터의 등장과 함께 IT 기반의 융·복합이 대세로 떠오르면서, 과거에는 존재하지 않았던 수많은 직업들이 탄생하고 있고, 동시에 많은 직업들이 사라지고 있다. 이렇게 탄생된 다수의 직업은 기업에 정착하기도 전에 세분화되거나 변형되어 또 다른 직업을 탄생시키는 경우도 허다하다. 급변하는 환경에 민첩하게 대처하기를 원하는 기업들에게는 이처럼 역동적으로 변화하는 직업 구조를 이해하기 위한 노력은 매우 중요할 것이다.

과거에는 직무분석을 통하여 기업에서 필요

로 하는 새로운 직무를 제안하고, 그 직무 수행에 필요한 지식을 정의해 왔다[6]. 일반적으로 업무에 필요한 지식은 하드 스킬(Hard Skill)과 소프트 스킬(Soft Skill)로 분류할 수 있다. 하드 스킬은 기술적 능력이나 전문적 지식을 의미하며, 소프트 스킬은 변화에 적응하는 능력, 커뮤니케이션, 협상, 팀워크, 리더십 등과 같이 책으로 단기간에 습득할 수 없는 개인 능력이라고 정의하고 있다[34]. 하드 스킬은 소프트 스킬에 비하여 문서화가 가능하고, 성과 측정이 쉽기 때문에 과거의 직무분석은 하드 스킬에 초점이 맞추어져 있었다[34]. 기업들은 직무분석을 통해 정의된 하드 스킬을 이용하여 인재를 선발하기 위한 기준으로 이용하였고, 그 이외에도 선발된 인재를 적절하게 배치하고, 그들을 평가하는 데도 널리 사용되어 왔다. 하지만 최근 빠르게 변화하는 경영환경에 있어서 전통적인 직무분석은 오늘날 조직에서 요구하는 빠른 외부 환경 변화에 적절하게 대응하지 못한다는 비판이 제기되고 있다[32]. Singh[33]은 전통적인 직무분

석 방법은 일정기간 지속적으로 수행되어지고 있는 직무에 초점을 맞추고, 무엇보다도 한 명의 종업원이 하나의 직무를 수행한다는 one-person, one-job 상황을 가정하였기 때문에 직무의 영역이 사전에 정의되어져 있어 지속적으로 변화하고 있는 상황에는 적합하지 않다고 주장하였다. 역동적으로 변화하는 환경에 대처하기 위하여 갑자기 새로운 직무가 탄생하기도 하고, 기존의 직무가 사라지거나 통합 혹은 분리되는 경우도 비일비재하기 때문에 전통적인 직무분석 방법으로는 새로운 직무를 능동적으로 디자인하기가 불가능할 것이다[15].

최근에 기업의 환경변화에 가장 크게 영향을 미친 요인 중에 하나는 빅데이터의 등장일 것이다. 빅데이터를 활용하여 과거에 기업이 수행하지 못하였던 전략을 이제는 빅데이터 분석을 기반으로 실시할 수 있게 됨으로써 빅데이터에 대한 가치를 기업들이 인지하게 되었다. 빅데이터의 역할과 가치의 중요성이 날로 높아지고 있는 4차 산업혁명 시대에 빅데이터와 관련된 직종은 신생 직종으로서 무한한 가능성을 가진 블루오션으로 자리매김하였다. 과거에는 분석이 가능하지 않았던 빅데이터를 분석할 수 있게 된 기업들은 이를 활용하여 새로운 기회를 창출하고자 노력하고 있다. 이런 기업의 노력의 중앙에는 데이터 과학자의 역할이 중요시되고 있다 [10]. 특히 인공지능, 사물인터넷 등 기술의 발전과 다양한 업무의 자동화로 ‘데이터 과학자’에 대한 기대와 요구는 계속해서 높아지고 있지만 기업에서 요구하는 적합한 기술력을 갖춘 핵심 전문인력의 공급은 수요에 비해 여전히 못 미치고 있는 현황이다. 이러한 문제를 초래한 원인으로서는 데이터 과학자의 직무에 대한 체계적인 연구가 충분하지 않아서, 일을 찾는 구직자와

일할 사람을 찾는 기업이 데이터 과학자에 대해서도 합의된 정의를 가지고 있지 못하다는 점을 주요 원인으로 들 수 있을 것이다[13]. 이에 데이터 과학자의 직무에 대한 이해를 돕기 위한 많은 연구가 현재 진행되어지고 있으나, 현재로는 직무의 영역이 명확하게 정의되어져 있지 않아서 전문가들마다, 기업마다 서로 다른 의견을 가지고 있어 대부분의 연구에서는 정량적인 데이터가 아닌 정성적인 데이터를 이용하여 각기 나름대로 데이터 과학자의 직무를 정의하고 있다[2, 28]. 특히 정성적인 데이터 중에서도 데이터 과학자에 대한 기업의 니즈를 효율적으로 알 수 있는 구직정보를 이용하여 데이터 과학자의 직무를 이해하고자 하는 노력이 최근에 많이 발견되어지고 있다[8, 9, 40]. 이런 트렌드의 연장 선 상에서 본 연구에서는 미국의 인터넷 구직사이트인 Glassdoor.com에서 데이터 과학자를 채용하기 위한 기업의 온라인 채용정보를 정성적으로 분석하여 실무현장에서 데이터 과학자에게 요구되는 역량이 무엇인지를 탐색하여 보고자 한다. 따라서 본 연구의 구체적인 연구 질문은 다음과 같다

PI: 데이터 과학자와 연관된 채용정보에서 가장 많이 등장하는 키워드는 어떤 것이 있는가?

PIa: 데이터 과학자에게 요구되는 교육 수준은 어느 정도인가?

PIb: 데이터 과학자가 익숙하게 이용할 줄 알아야 하는 분석 Tool과 프로그래밍 언어는 무엇인가?

PIc: 데이터 과학자와 연관된 채용정보에서 어떤 단어들(동시에 많이 등장하는가(N-Gram))?

P2: 데이터 과학자에게 요구되는 지식(하드 스킬과 소프트 스킬)은 무엇인가?

P3: 데이터 과학자에게 요구되는 소프트 스킬은 어떤 것 들이 있는가?

2. 빅데이터와 데이터 과학자

컴퓨터의 도입으로 데이터의 저장과 재생산은 전례 없는 확장을 가져왔다. 인터넷의 출현은 이러한 확장을 더욱 가속화 시켰고, 전세계 컴퓨터들의 상호연결은 데이터 생성의 폭발적인 증가를 야기했다. 데이터가 급속도로 확장된 것은 페이스북, 트위터, 유튜브 등과 같은 소셜 네트워크 서비스를 이용하여 사용자들이 직접 생성한 다양한 콘텐츠와 사용자들 간의 활발한 상호작용, 그리고 전자상거래 등과 같은 우리들의 온라인 활동이 큰 역할을 했다고 볼 수 있다[18]. 클라우드 컴퓨팅, 사물 인터넷과 같은 새로운 기술들의 출현은 이러한 추세를 더욱더 촉진시켰고, 특히 데이터가 폭발적으로 증가한 2000년대 후반부터는 빅데이터가 정보기술의 핵심 키워드로 자리 잡게 되었다.

Manyika et al.[21]은 빅데이터를 일반적인 데이터베이스 소프트웨어 관리툴(DBMS)로는 저장·관리·분석이 불가능할 정도로 큰 데이터 집합이라고 정의하였고, 빅데이터의 공통적인 특징은 일반적으로 3V(Volume, Variety, Velocity)로 설명할 수 있다고 정의하였다. 방대한 양(Volume), 다양한(Variety) 형태의 데이터, 그리고 데이터의 빠른 생성속도(Velocity)가 빅데이터의 기본적인 특징이라고 제시하였다. 빅데이터는 기하급수적으로 증가하고 있고, 기술의 발전은 이러한 데이터를 좀 더 쉽게 저장하고

처리하도록 돕고 있다. 빅데이터의 전략적 활용 여부가 기업의 경쟁력을 주도하게 되면서 선진국을 중심으로 빅데이터 기술과 활용에 대한 관심이 대폭 증가하고 있다. 데이터는 그 어느 때보다도 풍부하고 쉽게 얻을 수 있게 되었으며 기업들은 이러한 방대한 데이터 속에서 효율적인 정보를 찾기 위해 부단히 노력하고 있다. 이를 위해 기업 내·외부의 데이터를 효율적으로 수집하고, 정리하고, 분석하고, 활용할 수 있는 능력이 이제는 기업 역량의 중요한 척도가 되었다. 이와 더불어 데이터를 효과적으로 다룰 전문 인력인 데이터 과학자에 대한 관심도 증폭되었다[25].

데이터 과학자라는 용어는 기업의 방대한 양의 데이터를 구성하고 분석하는데 능숙한 데이터 전문가의 필요성을 깨닫게 되면서 새롭게 탄생된 직업이다. 지난 10년 간 데이터 과학자는 기업에 꼭 필요한 자산이 되었고, 거의 모든 조직에 서로 다른 직함으로 존재하고 있다. 그만큼 데이터 과학자에 대한 관심은 꾸준히 상승했고, 최근 몇 년 사이에는 폭발적인 상승 추세를 보였다. 데이터 과학자는 2012년 HBR(Harvard Business Review)에서 21세기 최고의 직업으로 소개되었을 뿐만 아니라[5] 미국 구인구직 플랫폼인 Glassdoor에서 발표한 25 Best Jobs in America에서 2016년부터 4년 연속 1위를 차지하였다. 영국의 Joint Information Systems Committee에서는 데이터 과학자를 문제해결을 위하여 관련된 사람들과 긴밀히 협력하여 창의적인 조사 및 분석에 참여할 수 있는 사람으로 정의하였다[37]. 가트너 그룹은 데이터 과학자를 복잡한 비즈니스 문제를 모델링하고, 통계학, 데이터 마이닝, 시각화 기법 등을 이용하여 방대한 양의 데이터 속에서 해답을 찾는 사람이라고

정의하였다. Manyika et al.[21]은 데이터 과학자를 빅데이터에 대한 이론적 지식과 분석 기술에 대한 숙련을 바탕으로 통찰력, 전달력, 협동 능력을 발휘할 수 있는 전문인력으로 정의하였다. Dhar[7]는 통계학, 기계학습, 인공지능, 데이터베이스 등을 포괄하는 종합적인 기술을 갖추고 문제 설계에 대한 이해가 깊은 사람으로 정의했다. Cho[4]는 데이터를 수집, 정리, 조사, 분석, 가시화할 수 있는 전문인력이라고 정의하였다. 이처럼 데이터 과학자는 데이터 분석 방법과 기술에 대한 지식과 스킬뿐만 아니라 문제를 이해하고, 모델링하는 능력, 더 나아가서 문제 해결을 위하여 다른 사람들과 효율적으로 협업하는 능력을 보유한 사람을 지칭하고 있다. 이와 같이 데이터 과학자에 대하여 너무나 폭넓게 정의되어 있어 데이터 과학자에게 필요한 지식을 세부적으로 정의하기가 힘든 상황이다.

3. 직무분석/직무역량

시대가 변함에 따라, 산업의 발전이나 다른 여러 가지 요인에 의해 어떠한 형태로든 직무는 계속해서 변화되고 있다. 따라서 해당 직무에 대한 내용이나 범주 또한 거듭 변화하고 있다. 전통적인 직무분석이 직무 단위로 세부적인 직무 내용과 자격 요건을 분석했다면, 현대적 직무 분석은 역량기반 접근법(Competency-Based Approach)을 도입하고 있다[31]. 현 시대는 직무의 표준화된 임무나 과업보다는 각각의 직무에서 우수한 능력을 나타내는 사람들의 특성을 더욱 중요하게 생각한다. 따라서 수행자의 능력을 도출하는 활동에만 급급한 단편적인 직무분석 방법보다 장기적이고 포괄적인 방법에 대한 고려

또한 필요할 것이다. 더욱이 직무 내용이 빠르게 변화하는 현대사회에서 전통적인 지식, 스킬, 능력에 대한 분석과 더불어 분석의 단위를 넓은 역량에 기반한 분석의 필요성이 증가되었다.

3.1 직무분석

직무분석에 대한 개념은 인사관리, 직업교육, 산업교육의 분야에서 다양하게 사용되어졌으며 특히 아담 스미스(A. Smith)의 분업론으로부터 유래되었다[26]. McCormick[23]에 따르면 직무분석은 직무의 특성에 관련된 여러가지 중요한 정보를 수집·분석·종합하고, 그 자료를 기반으로 각 직무에서 요구되는 지식과 기술, 특징, 자격요건 등 타 직무와 구별되는 요인을 분명하게 정의하여 업무에 대한 상세한 정보를 기술하는 체계적인 과정이라고 정의하고 있다[19]. 직무분석을 통하여 특정 직무를 구성하고 있는 세부적인 업무 내용과 직무를 성공적으로 수행하기 위하여 담당자에게 요구되어지는 경험, 자질, 기능, 지식, 능력 등을 세부적으로 정의할 수 있게 되었다[19]. 직무분석의 내용을 기반으로 그 직무에 대한 채용기준으로 사용하기도 하고 성과측정에도 널리 이용되고 있다.

직무분석 결과의 질을 좌우하는 것은 직무에 대한 정보의 정확성과 완전성이다. 직무에 대한 분석을 진행함에 있어서 직무와 관련된 정보를 얻는 가장 중요한 출처는 직무관련 전문가이다. 직무관련 전문가가 어떤 사람인지에 관한 자격요건이 정확하게 정해져 있지 않지만 최소 요건으로서 해당 직무에서 수행되는 모든 과제들에 대해 잘 알고 있어야 한다[38]. 직무관련 정보를 얻는데 가장 보편적으로 사용

되는 출처는 현재 직무에 종사하고 있는 사람 즉 현직자일 것이다. 직무관련 전문가로서 현직자를 사용하는 이유는 그들이 자신의 직무에 대해 가장 상세하게 잘 알고 있기 때문이다. 현직자들 중에서도 자신이 하고 있는 일에 대해 잘 알고 있고 그것을 말로 잘 표현할 수 있는 사람이 있는 반면 상대적으로 그러한 능력이 부족한 경우도 있을 수 있다. 따라서 현직자들 중에서도 어떤 사람을 선정하는지가 직무분석 결과의 질이 좌우될 수 있을 것이다[16]. 현직자들의 언어능력, 기억력, 협조성과 같은 개인적 특성들도 그들이 제공하는 정보의 질을 크게 좌우할 것이다. Sanchez and Levine[30]은 과거에는 직무분석을 하기 위해 정보를 얻는 과정에서 현직자에게 가장 많이 의존했지만 앞으로는 다른 곳으로부터 직무에 관한 정보를 얻을 필요성이 증가될 것이라고 하였다. 데이터 과학자와 같이 기존에 존재하였던 직무가 아니라 새롭게 등장하였고 직무의 범위 또한 지속적으로 변화하고 있는 시점에서 데이터 과학자에게 요구되는 지식과 능력을 현직자에게 의존한 전통적인 방법으로 정의하는 것은 매우 비효율적인 방법이라고 생각된다.

전통적인 직무분석은 특정직무와 관련하여 수행되는 구체적인 과업과 활동들에 대해 세부적인 정보를 제공해 준다. 이러한 정보는 직무들 간의 경계를 설정하고 각 직무를 수행함에 있어서 개인에게 과제와 책임을 부여하는데 사용되었다. 4차 산업혁명 시대를 맞아 최근 조직이 처한 환경은 급격한 변화를 가져옴과 동시에 새로운 기술들이 끊임없이 개발되고 있다. 따라서 개인이 담당해야 할 직무의 범위와 역할은 과거보다 훨씬 빠른 변화와 확장을 맞이하였다. 이로 인한 직무의 내용은 과거에 비해

전반적으로 빠르게 변하고 있고, 직무가 빠르게 변할수록 기존에 얻었던 직무관련 정보의 유용성은 감소되기 마련이다. 조직 내에서 개인들이 수행하는 업무에 대한 분석은 여전히 필요하고 중요하지만 과거의 직무분석 방법을 그대로 따르는 것 보다는 이러한 변화에 맞춰 직무분석의 방법이나 형태도 또한 변화되어야 할 것이다. Sanchez and Levine[30]은 21세기의 직무에는 많은 변화가 있을 것이므로 이러한 변화에 발맞춰 과거에 행하였던 전통적인 직무분석의 방식 또한 변화되어야 한다고 주장하면서 전통적인 방법으로 수집하였던 지식, 기술, 능력보다는 분석 단위를 넓혀 역량 중심으로 분석할 필요가 있음을 강조하였다. 역량은 유사한 지식과 기술 및 능력들을 묶어서 보다 큰 단위로서 인적 특성을 정의한 것이다. 직무에 대해 지나치게 세분화하여 정보를 얻게 되면 직무가 조금만 변하여도 금방 쓸모없게 될 수도 있지만, 보다 큰 단위로 정리된 결과는 직무가 변화되어도 그대로 적용될 가능성이 높기 때문이다. 과거의 직무분석은 기업의 입장에서 직무를 정의하였다고 하면, 역량을 기반으로 한 직무분석은 직무를 직접 수행하는 종업원의 관점에서 직무를 정의하였다고 볼 수 있을 것이다[32].

3.2 직무역량

역량이란 개념은 1970년대 하버드 대학 McClelland 교수가 전통적인 학업, 적성 검사 혹은 성취도 검사의 한계를 지적하면서 사용하였다. McClelland[22]는 인재를 선발하고 훈련시키는 데 있어서 보다 효과적인 방법은 직무 특성에 기초한 역량을 활용하는 것이라고 하였다. 1980

<Table 1> Job Competencies of IT Employees

	Competencies
Skill (A learned ability to perform a task)	1. Critical Analysis & Problem-Solving Skills 2. Communication Skills 3. Conceptualization and Abstraction Skills 4. Skills to Manage Situational Politics
Knowledge (Acquired information in specific domain)	1. Technical Knowledge 2. Work Experiential Knowledge 3. Comprehensive Knowledge 4. Contextual Knowledge
Self-Concept (Attitudes, Values, or self-images)	1. Work the Middle Ground(Link between IT and the rest of business) 2. Visionary
Traits (Dispositional characteristics)	1. Be Creative 2. Be Open-minded
Motives (Internal drives & need to seek achievement)	1. Be Passionate 2. Be Resilient(Conflict Management)

Source: Ho and Frampton[16].

년대에 들어서면서 미국의 기업들은 효율적인 관리에 많은 관심을 기울이게 되었고, 학자들은 관리자의 역량에 대해 연구하기 시작하였다. Boyatzis[1]는 역량에 대해 업무를 수행함에 있어서 우수한 성과를 나타내는 개인의 내재적 특성이라고 정의하고, 이는 개인이 가지고 있는 지식, 기술, 특질, 동기, 자기 이미지 및 사회적 역할의 종합체라고 정의하였다. Saif et al.[29]은 직무역량을 직무를 수행하는데 있어서 기업으로부터 요구되는 지식, 기술, 경험, 능력 등을 포괄적으로 지칭하는 것으로, 전통적인 직무분석에서 중요시하는 단기적인 평가보다는 조직이나 개인의 우수한 능력을 발굴하는 장기적인 목표에 초점을 둔 방법이라고 정의하였다. Ho and Frampton[12]은 IT 분야 종사자들의 직무역량을 전문가들과의 인터뷰를 통하여 <Table 1>과 같이 정의하였는데, 여기에는 지식(Knowledge)이나 기술(Skill) 뿐만 아니라 태도(Self-Concept)나 동기(Motive), 심지어는 개인적인 속성(Trait)까지도 포함된다.

3.3 데이터 과학자 역량

빅데이터 시대를 맞으면서 데이터 과학자의 역량에 대한 연구는 활발히 진행되어 왔다. 각각의 연구들은 나름대로의 관점에서 데이터 과학자의 역량에 대해 다양하게 정의하고 있다. 그리고 계속해서 발전하고 있는 빅데이터 기술로 인해 데이터 과학자 역량에 대한 정의 또한 지속적으로 진화하며 갱신되고 있다.

LaValle et al.[17]은 IBM 기업가치연구소와 제휴하여 30개 이상의 산업 및 100개국에서 약 3,000명의 전문가를 대상으로 설문조사를 진행하여 7가지의 데이터 과학자가 가지고 있어야 할 역량을 제시하였다: (1) 데이터 시각화, (2) 시뮬레이션(simulation) 및 시나리오 개발, (3) 비즈니스 프로세스 내에 분석 적용, (4) 회귀분석, 이산 선택 모형, 수학적 최적화, (5) 과거의 추세 분석 및 미래 동향 예측, (6) 클러스터링 형성 및 세분화, 그리고 (7) 표준화된 보고 능

력. Patil[27]은 데이터 과학자는 일정 과학분야에 대한 전문지식과 기술적 숙련도가 필요하고 호기심과 문제해결 능력, 효과적인 대화를 위한 스토리텔링과 창의적으로 문제를 바라보는 능력이 필요하다고 하였다. Kart and Laney[14]는 데이터 과학자의 역량에 대해 데이터 관리, 분석 모델링, 비즈니스 분석, 그리고 커뮤니케이션, 협력, 리더십, 창의성 및 열정 등 소프트 스킬로 정리하였다. National Information Society Agency[20]에서는 데이터 과학자의 역량에 대해 (1) 통계, 데이터 마이닝, 기계학습, 자연어 처리 등 빅데이터에 대한 이론적 지식, (2) 분석 기술에 대한 숙련, (3) 창의적 사고, 호기심, 논리적 비판을 탑재한 분석 통찰력, (4) 스토리텔링, 시각화를 겸비한 설득력 있는 전달, (5) 다분야간 협력을 위한 커뮤니케이션 등 5가지로 정리하였다. Cho[3]은 데이터 과학자 역량을 (1) 정보기술, (2) 수학 및 통계기술, (3) 현업 업무지식, (4) 호기심, 창의성, 객관성, 논리적 사고, 인내심, 상식 등의 소프트 스킬로 구분했고, Metcalf and Brenza[24]는 데이터 과학자에게 요구되는 필수적인 숙련기술을 (1) 분석역량(수학, 정보기술, 비즈니스 지식), (2) 의사소통, (3) 호기심, (4) 협동, (5) 비즈니스 감각, (6) 고객중심적 사고, (7) 문제해결능력, (8) 선제적으로 주도하는 자세, (9) 전략적 사고, (10) 노력과 열정 등 10가지를 제시하였다. 살펴본 여러 문헌에서 정의한 데이터 과학자의 역량은 서로 다른 면을 강조하고 있으나, 공통적으로 데이터 과학자는 기술에 대한 이해, 비즈니스에 대한 이해, 문제해결 능력, 더 나아가서 효율적인 팀워크를 위한 기본적인 소양이 필수적으로 요구됨을 알 수 있었다.

4. 연구설계 및 조사방법

본 논문에서는 인터넷 구직 사이트에서의 온라인 채용정보를 이용하여 데이터 과학자의 역량을 탐색하는데 주요 목적을 두고 있다. 채용정보를 분석하기 위하여 텍스트 분석도구 중 하나인 NVivo를 활용하였다. NVivo는 전 세계적으로 학자 및 전문 연구자가 가장 많이 사용하는 정성데이터 분석 소프트웨어로 텍스트, 이미지, 오디오 및 비디오, 데이터 세트 등 다양한 데이터 소스를 관리하고 분석하는 기능을 폭넓게 제공하고 있다. NVivo를 사용하여 분석하기 위해 가장 먼저 수집한 자료를 분석 프레임워크에 맞추어 분류한 후, 분류된 키워드 그룹을 기반으로 다양한 분석을 실시하였다.

본 연구에서는 데이터 과학자에 대한 채용정보를 수집하기 위해서 세계에서 가장 큰 채용 사이트인 Glassdoor 사이트를 이용하였다. Glassdoor 사이트에서 data scientist라는 키워드로 채용정보를 검색하였다. 2020년 1월과 10월 두 차례에 걸쳐 각각 2,120개와 2,130개의 자료를 수집하여, 총 4,000여 개의 자료를 분석에 활용하였다.

5. 분석

수집된 채용정보를 NVivo를 이용하여 다음과 같은 분석을 실시하였다.

5.1 키워드 빈도분석

P1: 어떤 키워드가 데이터 과학자 채용정보에 가장 많이 등장하는가?

데이터 과학자의 주요 역량 항목을 도출하기 위해 1차 분석으로 단순 단어 빈도분석을 실시하여 상위 500개를 도출한 후, 도출된 500개의 키워드 중 의미 있는 키워드를 다시 추출한 후 유사한 의미를 가진 키워드들은 동일한 노드로 재분류하였다. <Table 2>는 재 분류된 노드 중에서 상위 20개 노드의 빈도분석 결과를 보여준다. 데이터(Data)라는 키워드가 빈도 상으로는 압도적으로 많이 등장하였지만, 채용정보의 수를 기준으로 보았을 때에는 97.8%의 채용공고에서 언급된 경험(Experience)라는 키워드가 더 많이 등장하였음을 알 수 있었다. 그 뒤로는 일(Work), 개발(Development), 스킬(Skill), 그리고 커뮤니케이션(Communication) 등의

순서로 등장 빈도가 높았고, Python은 프로그래밍 언어 중에서 유일하게 순위에 올랐다. 높은 빈도를 나타낸 상위의 키워드들을 포괄적으로 살펴보았을 때, 데이터 과학자는 데이터와 밀접한 연관이 있는 업무를 처리하는 사람으로 관련 직무에 대한 경험과 스킬이 요구되고, 커뮤니케이션 스킬이 매우 중요하게 작용하는 직무를 알 수 있었다. 키워드 빈도 분석을 통해 알아낸 데이터 과학자 직무에 대한 특징은 데이터 과학자의 직무를 정의하는데 목적을 둔 여러 연구에서 주요하게 언급한 항목과 일치함을 발견하였다[9, 11, 22].

<Table 2> Keyword Frequency Analysis(Top 20)

Keyword	No. of Frequency	% of Ads
Experience	11955	97.8%
Data	22021	97.4%
Work	9028	94.1%
Development	7267	85.8%
Skill	4439	79.2%
Communication	3401	78.5%
Science	4567	74.0%
Analytics	5751	69.4%
Management	4025	66.2%
Process	3310	65.5%
Learns	4451	64.0%
Analysis	3561	63.8%
Model	4686	63.4%
Technology	2985	60.8%
Information	2916	60.5%
Environment	2265	59.5%
Engineering	3405	59.5%
Knowledge	2534	59.2%
Python	1688	59.1%

PIa: 데이터 과학자에게 요구되는 교육 수준은 어느 정도 인가?

데이터 과학자에게 요구되는 교육수준에 대한 분석결과는 <Table 3>에 요약되어 있다. 87.6%의 채용정보에서 교육에 대해 언급하였고, 그 중에서 732건(55%)이 학사학위를, 689건(51.8%)이 석사학위를, 그리고 445건(33.5%)가 박사학위를 요구하는 것으로 나타났다. 이는 데이터 과학자에 대한 요구가 고학력을 선호했던 과거와는 달리 현장에서의 실무경험을 선호하는 쪽으로 기울고 있음을 간접적으로 알 수 있었다.

<Table 3> Level of Education Required for Data Scientists

Degree	No. of ads	Percentage
Bachelor	732	55.0%
Master	689	51.8%
PhD	445	33.5%
Total	1330	100.0%

PIb: 데이터 과학자가 능숙하게 다루어야 할 Tool과 프로그래밍 언어는 무엇인가?

<Table 4>는 상위 15개의 데이터 과학자에게 요구되는 Tool에 대한 결과를 정리하였다. Python이 전체 채용정보의 59.1%에 언급되어 가장 많았고, SQL이 50.1%로 두 번째로 많았다. Kim and Lee[18]의 연구에서는 R이 가장 많았지만 본 연구에서는 3위에 머물렀다. R은 주로 통계분석에 사용되고 Python은 데이터 과학에 대한 보다 일반적인 접근방식을 제공하고 있어, 시간이 지나면서 기업에서 관심이 증가하고 있음을 간접적으로 알 수 있었다. 전반적으로 데이터 과학자가 능숙하게 다루어야 할 Tool에는 프로그래밍 언어, 통계 패키지, 데이터베이스 도구, 빅데이터 플랫폼 및 사무용 애플리케이션 등이 다양하게 포함되어 있음을 알 수 있었다.

<Table 4> Data Scientist Tools by Frequency

Keyword	# of Freq.	% of Ads
Python	1688	59.1%
SQL	1593	50.1%
R	1279	42.9%
Tableau	551	19.7%
Spark	603	19.1%
Excel	484	18.7%
C	575	17.1%
Java	416	16.8%
Hadoop	348	16.3%
AWS	319	15.0%
SAS	263	12.3%
Tensorflow	201	9.4%
Hive	200	9.4%
Azure	183	8.6%
Linux	144	6.8%

PIc: 채용정보에서 어떠한 단어들이 연속적으로 나열되어 있는가(N-Gram)?

데이터 과학자에게 요구되는 역량을 탐색하기 위하여, 앞에서 실시한 단일 키워드에 대한 빈도 분석뿐만 아니라 2개, 3개의 키워드가 연결되어 동시에 많이 등장하는 키워드들도 탐색하였다. NVivo의 Word Tree 분석을 이용하여 두 개의 키워드(Bigram) 혹은 세 개의 키워드(Trigram)가 함께 등장하는 빈도 수를 탐색하는 N-Gram 분석을 실시하였다. <Table 5>은 채용정보에 자주 등장하는 연속적인 단어의 나열을 보여주는 상위 7개의 N-Gram 분석결과이다. 두 개의 키워드가 함께 등장한 빈도가 많은 키워드들을 분석한 Bigram 분석 결과를 기반으로, 기계학습(Machine Learning), 전산학(Computer Science), 데이터 과학(Data Science) 등과 같은 특정 분야에 대한 지식이나 기술뿐만 아니라 그 분야에 있어서의 다년간의 경험(Years of Experience) 또한 기업에서 요구하는 것으로 유추할 수 있었다. 또한 커뮤니케이션 기술(Communication Skill) 등도 데이터 과학자에게 요구되는 주요한 역량임을 알 수 있었다. 세 개의 키워드가 함께 등장한 키워드들의 빈도를 분석한 Trigram 분석에서는 기계학습과 연관된 다양한 키워드들(Machine Learning Techniques, Machine Learning Model, Machine Learning Algorithms, Natural Language Processing, Large Scale Data 등)이 상위에 위치한 것으로 미루어 보아 기계학습에 대한 폭넓은 지식, 기술, 그리고 경험이 동시에 요구됨을 알 수 있었다. 또한 문제 해결 능력(Problem Solving Skill)도 요구됨을 알 수 있었다. Bigram 분석으로부터 도출한 커뮤니케이션 기술을 Trigram

분석을 통하여 문서작성 스킬(Written Communication Skill)과 발표 기술(Verbal Communication Skill)로 구체화시킬 수 있었다.

<Table 5> N-Gram Analysis

Bigrams	# of Freq	% of Ads	Trigrams	# of Freq	% of Ads
Machine Learning	3134	46.1	Large Data Sets	295	11.7
Computer Science	1123	43.0	Written Comm. Skills	246	11.1
Data Science	2218	40.0	Problem Solving Skills	225	10.4
Years Exp.	1220	36.9	Verbal Comm. Skills	188	8.8
Comm. Skills	814	34.9	Machine Learning Tech.	188	7.6
Data Set	1210	34.6	Natural Language Processing	200	7.6
Data Analysis	929	29.0	Machine Learning Models	199	7.1

키워드 분석을 통하여 데이터 과학자는 데이터를 관리하고, 분석하는 지식, 즉 하드 스킬 뿐만 아니라 커뮤니케이션과 문제 해결 능력과 같은 소프트 스킬도 현재 기업에서 널리 요구됨을 알 수 있었다.

5.2 직무역량 분석 프레임워크

P2: 데이터 과학자에게 요구되는 지식은 무엇인가?

Kim and Lee[22]는 IS 관리자에게 요구되는 직무역량을 분류한 Todd[39]의 분석 프레임워크를 기반으로 데이터 과학자에 대한 분석을 진행하여 8개의 직무역량을 도출하였다. Kim and Lee[18]는 데이터 과학자에게 요구되는 역량을 먼저 시스템 스킬(System Skill), 비즈니스 스킬(Business Skill), 그리고 기술적 스킬(Technical Skill)로 대분류하고, 세부적으로 시스템 스킬을 개발(Development) 스킬, 문제해결(Problem Solving) 스킬로, 비즈니스 스킬을 사회성(Social), 비즈니스(Business), 그리고 관리(Management)와 연관된 스킬로, 기술적 스킬을 소프트웨어(Software), 네트워크(Architecture/Network), 그리고 하드웨어(Hardware)와 연관된 스킬로 재 정의하였다. 본 연구의 앞선 분석에서 도출한 주요 키워드들을 Kim and Lee[18]가 정의한 8개의 세부 직무역량을 기반으로 재 분류하여 <Table 6>와 같은 결과를 얻었다. Kim and Lee[18]가 정의한 8개의 중분류 스킬 중에서 기술적 스킬에 포함된 네트워크(3b)(57.1%)와 하드웨어(3c)(17.1%)에 대한 스킬을 제외한 다른 6개의 중분류 스킬은 75% 이상의 채용정보에 적어도 한 번 이상 언급되었음을 알 수 있었다. 이는 하드웨어에 대한 지식이나 스킬보다는 소프트웨어에 대한 지식이나 스킬이 데이터 과학자에게 더 요구됨을 알 수 있었다. 또한 기술과 분석 능력 뿐만 아니라 팀으로 작업하면서 의사결정자, 동료, 팀원 등 여러 이해관계자와의 소통 능력이 중요함을 알 수 있었다. 또한 데이터 분석 결과가 적용되어야 할 비즈니스 자체에 대한 지식이나 스킬(Business: 76%)보다는 데이터 기반 의사결정 프로세스를 전반적으로 관리할 수 있는 프로젝트 관리 지식(Management: 95.5%)이 상

<Table 6> Knowledge Framework

System Skill Class	% of Ads	Business Skill Class	% of Ads	Technical Skill Class	% of Ads
(1a) Development (Hard Skills)	98.6%	(2a) Social (Soft Skills)	96.8%	(3a) Software (Hard Skills)	97.1%
Technology	86.2%	Team	86.0%	python	59.1%
Analysis	79.6%	Communication	78.5%	SQL	50.1%
Design	58.8%	collaboration	54.1%	machine learning	46.1%
Programming	57.7%	group	28.5%	R	42.9%
Implementation/Testing	47.4%	independently	25.9%	database/ data warehouse	22.3%
Documentation	26.8%	Interpersonal	12.7%	applications	65.4%
Methodology	20.0%	Self-Motivated	5.0%	Data visualization/Data processing	19.5%
Data management prototype	9.9%	presentation	18.5%	Platforms	33.4%
				spark	19.1%
				Java	16.8%
				tableau	19.7%
		(2b) Business (Hard Skills)	76.2%	frameworks	16.7%
		services	48.1%	hadoop	16.3%
		Marketing/Sales	33.0%	excel	18.7%
		Finance	24.5%	tracking	18.4%
		accounting	15.3%	programming language	14.8%
		consulting	14.4%	data mining	16.7%
		Economic	13.8%	SAS	12.3%
				C	17.1%
				deep learning	11.6%
				hive	5.6%
				tensorflow	9.4%
				linux	9.4%
		(2c) Management (Soft Skills)	95.5%	matlab	6.8%
		responsible	68.8%	NoSQL	7.8%
		organization	50.2%	nlp	6.6%
		decisive	41.3%	(3b) Architecture/Network (Hard Skills)	57.1%
		planning	35.2%	cloud	22.3%
(1b) Problem Solving (Hard & Soft Skills)	94.2%	monitor & control	30.8%	security	21.1%
Analytical	69.4%	training	30.5%	networking	18.8%
modeling	63.4%	strategy	28.6%	Architecture	15.2%
Math/Statistics	62.3%	recommendations	26.2%	aws	15.0%
creative/innovative	51.4%	leadership	23.6%	web	11.6%
regression	11.2%	guidance	18.7%	azure(클라우드 컴퓨팅 플랫폼)	8.6%
troubleshooting	10.0%	dashboards	15.0%	(3c) Hardware(Hard Skills)	17.1%
forecasting	9.5%	propose	9.2%	storage/Devices	11.6%
mappings	6.1%	project management	8.4%	server	6.6%

대적으로 중요하게 여겨지고 있음을 알 수 있었다. 35명의 전문가 인터뷰를 기반으로 한 Yoon et al.[41]에 의하면 데이터 분석 역량(<Table 6>에서의 1a & 1b와 유사)이 가장 중요한 데이터 과학자 역량으로 전문가들은 꼽았다. 두 번째로 중요한 역량으로 비즈니스 역량(<Table 6>에서의 2b & 2c와 유사)을, 다음으로 기술 역량(<Table 6>에서의 3a, 3b & 3c와 유사)을, 마지막으로 사회화 역량(<Table 6>에서의 2a와 유사)이 상대적으로 덜 중요한 요인임을 확인하였다. 이 연구 결과는 본 연구의 결과와는 약간의 차이가 있음을 알 수 있었다. 물론 한국 기업과 미국 기업의 차이일 수도 있으나, 한편으로는 전문가들의 생각과 기업에서의 채용공고에 차이가 있음을 간접적으로 알 수가 있었다. 가장 크게 차이가 나는 부분은 본 연구에서는 사회화 역량(2a)이 많은 데이터 과학자의 채용공고에 등장하였으나, Yoon et al.[41]에서는 사회화 역량의 중요도가 낮게 평가된 점이다.

5.3 소프트 스킬

P3: 데이터 과학자에게 요구되는 소프트 스킬은 어떤 것 들이 있는가?

데이터 과학자에게 요구되는 소프트 스킬과 긴밀하게 연관이 되어진 주요 키워드들의 전·후 단어를 분석하면서 좀더 심도 있게 데이터 과학자에게 요구되는 소프트 스킬이 어떤 것들이 있는지를 탐색해 보았다. 본 연구에서는 앞에서 실시한 빈도분석에서 높은 순위에 언급되었던 키워드들 중에서 소프트 스킬과 긴밀하게 연관되어졌다고 판단되는 3개의 단어를 선택하여

추가적인 분석을 실시하였다. 본 연구에서는 소통능력(Communication), 리더십(Leadership), 그리고 창의성(Creativity)을 선택하고, 이 키워드들을 중심으로 전·후에 어떤 단어들 이 자주 등장하였는지를 탐색하는 Word Tree 분석을 실시하였다.

첫 번째는 소통능력에 관한 분석결과이다. 세부적으로 어떤 내용을 누구와 소통하는 능력이 데이터 과학자에게 요구되는지를 탐색하기 위해 <Figure 1>에 보는 바와 같이 소통능력이란 키워드 뒤에 어떤 키워드들이 함께 등장하는지를 탐색하여 보았다. 소통능력이라는 키워드가 등장한 대부분의 채용정보에 팀 활동과 연관된 키워드들과 함께 등장함을 발견하였다(<Figure 1>, A&C). 이것으로 미루어 보아 데이터 과학자에게는 효율적인 팀 활동을 위한 소프트 스킬들이 필요함을 유추할 수 있었다. 구체적으로 어떤 소프트 스킬이 필요한지를 살펴보면, 소통능력이 요구되는 데이터 과학자 채용정보 중에서 74% 이상의 채용정보가 어려운 기술적인 내용이나 분석 결과를 실무자나 경영자에게 효율적으로 설명할 수 있는 소통능력을 요구하고 있음을 알 수 있었다(<Figure 1>, B&D).

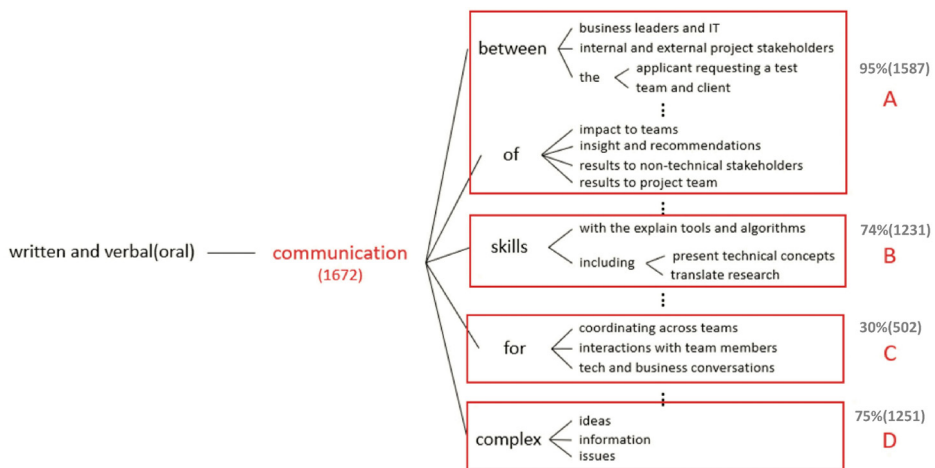
두 번째는 리더십에 관한 분석결과이다. 다양한 여러 이해관계자들과 함께 업무를 수행하기 위해서는 그들의 행동을 하나의 공동 목표를 향해 이끌어 나아가야 할 필요성이 있을 것이다. 이 과정에서 필요한 것은 데이터 과학자의 적극적인 리더십일 것이다. 리더십과 연관된 Word Tree 분석을 통해 데이터 과학자에게 요구되는 구체적인 리더십의 역할에 대해 탐색할 수 있었다. 리더십이 데이터 과학자에게 요구되는 채용정보 중에서 97%가 새로운 프로젝

트를 기획하고, 프로젝트 진행과정을 관리하는 프로젝트 관리와 연관된 리더십이 강조되고 있음을 발견하였다(<Figure 2>, C). 이외에 팀 내에서 원활하게 지식을 공유하기 위하여 여러 이해관계자들 간의 매개자 역할과 팀 운영에 있어서 가이드라인과 조언을 제공하여 줄 수 있는 Mentor 역할을 데이터 과학자가 해 줄 것을 기대하고 있는 것을 알 수 있었다(<Figure 2>, A). 이는 팀의 효율적인 운영을 위한 리더십이 데이터 과학자에게 요구됨을 알 수 있었다. 분석에 의하면 기술적인 의사결정뿐만 아니라 전략적 의사결정에도 주도적인 역할을 해 줄 것을 요구하고 있는 것으로 조사되었다(<Figure 2>, B). 이와 같은 결과를 바탕으로 데이터 과학자는 기술적인 지식뿐만 아니라 비즈니스와 연관된 지식을 폭넓게 보유하고 있어 기술자/엔지니어와 경영자의 가교 역할을 해 줄 것을 기업에서는 기대하고 있는 것으로 유추할 수 있을 것이다.

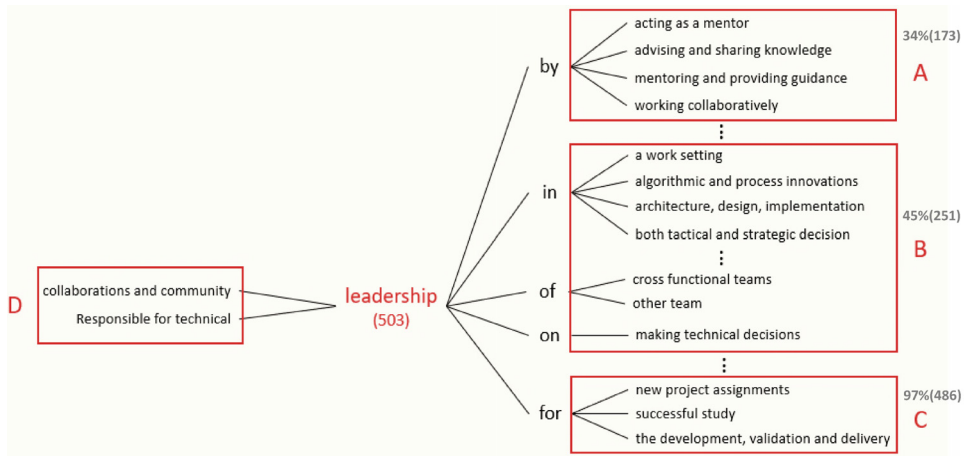
세 번째는 창의성에 관한 분석결과이다. 창

의성은 기존의 지식을 바탕으로 새로운 무언가를 창출해내는 사고의 능력이다. 데이터 과학자는 기존의 문제 해결 방식이 아닌 새로운 방법으로 문제를 해결할 수 있는 능력이 있어야 함을 <Figure 3>의 박스 A 분석 결과를 통해 알 수 있었다. 이는 기존의 문제 해결 방식과는 달리 이제는 빅데이터를 기반으로 기업의 문제를 창의적으로 해결할 수 있는 데이터 과학자가 필요해졌음을 알 수 있었다. 더불어 창의적으로 문제를 해결하려는 열정 또한 요구됨을 알 수 있었다. 당면한 문제에 대한 해결뿐만 아니라 좀 더 능동적으로 혁신적인 제품이나 서비스를 개발하여 변화를 추구하는데 주요한 역할을 해야 하는 Change Agent의 역할도 해야 함을 강조하고 있다(<Figure 3>, B&C).

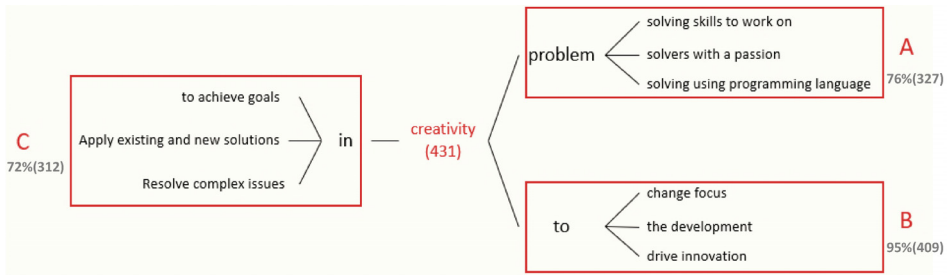
데이터 과학자에게 요구되는 소프트 스킬을 Word Tree 분석을 통해 살펴본 결과, 단순히 시스템을 개발하거나 데이터를 분석하는 것 이외의 추가적인 역량을 요구함을 알 수 있었다. 과거에 비즈니스 기획자나 컨설턴트에게 기대



<Figure 1> Word Tree Analysis for Communication



〈Figure 2〉 Word Tree Analysis for Leadership



〈Figure 3〉 Word Tree Analysis for Creativity

하였던 역할을 데이터 과학자에게 요구하고 있음을 알 수 있었다. 이는 기업 전반에 데이터 기반 의사결정이 중요해지면서 데이터를 관리·분석하는 기능 뿐만 아니라 데이터 분석 결과를 실제적으로 기업에서 활용할 수 있도록 하는 프로세스를 총괄적으로 관리하는 책임자 역할을 데이터 과학자에게 요구하고 있음을 알 수 있었다.

6. 결 론

본 연구에서는 온라인 채용정보를 정성적으

로 분석하여, 최근 대중의 많은 관심을 받고 있는 데이터 과학자에게 요구되는 직무역량을 탐색하였다. 그 결과, 과거 소프트웨어 엔지니어나 시스템 분석가들에게 요구되었던 Technical Skill과 System Skill 뿐만 아니라 비즈니스 컨설턴트나 Project Manager에게 요구되었던 비즈니스 관련 스킬이나 효율적인 팀워크를 위한 대인관련 스킬이 광범위하게 요구됨을 발견하였다. 과거에는 존재하지 않았던 직업들이 매일 새롭게 등장하고 있다. 이렇게 새롭게 등장한 직업들은 기존의 직무들이 환경의 변화에 맞추어서 새롭게 조합을 형성하면서 탄생되는 것으로 이해할 수 있을 것이다. 지속적으로 기업의

환경이 변화하고 기술이 발전되면서 데이터 과학자에게 요구되는 역량 역시 지금도 계속 진화하고 있다. 최근에는 관리자와 기획자의 역할이 중요시되고 있는 데이터 과학자와 구분하여 데이터 수집·관리·분석과 같은 기능적이고 기술적인 직무에 초점을 맞춘 Data Curators[36, 37] 혹은 Data Engineers[35]라는 새로운 직종이 소개되어지고 있다. 이런 역동적인 변화를 이해하기 위해서는 현재 기업의 니즈를 가장 잘 표현해 줄 수 있는 채용공고를 체계적으로 분석해야 하는 노력이 필요할 것이다. 이에 본 연구에서는 채용공고에 제시된 내용을 기반으로 직무역량을 도출하기 위한 새로운 방법론을 제시하였다는데서 그 의의를 찾을 수 있을 것이다.

본 연구는 다음과 같은 한계점을 가지고 있다. 첫째, 본 연구에서 정량적인 분석이 아니라 정성적인 분석 방법론을 사용하였기 때문에 주관성이 다소 포함되어 본 연구의 결과를 보편화시키는데 한계가 있을 것으로 판단된다. 둘째, 데이터 과학자에게 필요한 역량은 직급(Junior, Senior, Principal)에 따라 다를 것으로 생각한다. 그러나 본 연구에서는 직급을 고려하지 않고 전체 데이터 과학자의 역량에 초점을 맞추어서 연구를 진행하였다. 본 연구의 분석 결과는 향후 설문조사와 같은 계량적인 방법으로 검증할 필요성이 있고 데이터 과학자의 레벨에 따른 각각의 역량에 대해서 파악할 필요가 있다.

References

- [1] Boyatzis, R.E., *The competent manager: A model for effective performance*, New York: Wiley, 1982.
- [2] Chen, C. and Jiang, H., "Important skills for data scientists in china: two delphi studie," *Journal of Computer Information Systems*, Vol. 60, No. 3, pp. 287-296, 2018.
- [3] Cho, W.S., "Big Data Era, Data Scientist Training Plan," *Science and Technology Policy*, Vol. 192, pp. 44-55, 2013.
- [4] Cho, W.S., "Big Data Utilization and Data Scientist," *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 32, No. 1, pp. 59-65, 2014.
- [5] Davenport, T.H. and Patil, D.J., "Data scientist," *Harvard business review*, Vol. 90, No. 5, pp. 70-76, 2012.
- [6] DeAngelis, J.T. and Wolcott, M.D., "A job analysis to define the role of the pharmacy preceptor," *American journal of pharmaceutical education*, Vol. 83, No. 7, pp. 1480-1491, 2019.
- [7] Dhar, V., "Data science and prediction," *Communications of the ACM*, Vol. 56, No. 12, pp. 64-73, 2013.
- [8] Ecleo, J.J. and Galido, A., "Surveying LinkedIn profiles of data scientists: The case of the Philippines," *Procedia Computer Science*, Vol. 124, pp. 53-60, 2017.
- [9] Gardiner, A., Aasheim, C., Rutner, P., and Williams, S., "Skill requirements in big data: A content analysis of job advertisements," *Journal of Computer Information Systems*, Vol. 58, No. 4, pp. 374-384, 2018.
- [10] Gehl, R.W., "Sharing, knowledge man-

- agement and big data: A partial genealogy of the data scientist,” *European journal of cultural studies*, Vol. 18, No. 4-5, pp. 413-428, 2015.
- [11] Hattingh, M., Marshall, L., Holmner, M., and Naidoo, R., “Data Science Competency in Organisations: A Systematic Review and Unified Model,” *Proceedings of the South African Institute of Computer Scientists and Information Technologists*, pp. 1-8, 2019.
- [12] Ho, S.Y. and Frampton, K., “A competency model for the information technology workforce: Implications for training and selection,” *Communications of the Association for Information Systems*, Vol. 27, No. 1, pp. 1-21, 2010.
- [13] Hu, H., Luo, Y., Wen, Y., Ong, Y. S., and Zhang, X., “How to find a perfect data scientist: A distance-metric learning approach,” *IEEE Access*, Vol. 6, pp. 60380-60395, 2018.
- [14] Kart, L. and Laney, D., “Emerging Role of the Data Scientist and the Art of Data Science,” March, Gartner Research, <https://www.gartner.com/en/documents/1955615/emerging-role-of-the-data-scientist-and-the-art-of-data->, 2012.
- [15] Landau, K. and Rohmert, W. (Eds.), *Recent developments in job analysis*, Taylor & Francis, 2017.
- [16] Landy, E.J. and Vasey, J., “Job analysis: The Composition of SME Samples,” *Personnel Psychology*, Vol. 44, pp. 27 - 50, 1991.
- [17] LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., and Kruschwitz, N., “Big data, analytics and the path from insights to value,” *Sloan Management Review*, Vol. 52, No. 2, pp. 21-32, 2011.
- [18] Kim, J.Y. and Lee, C.K., “An empirical analysis of requirements for data scientists using online job postings,” *International Journal of Software Engineering and Its Applications*, Vol. 10, No. 4, pp. 161-172, 2016.
- [19] Kim, J.H., “Real estate training course development through job analysis,” *Korea Real Estate Academy Review*, Vol. 59, pp. 231-245, 2014. (in Korean)
- [20] National Information Society Agency, “Human Resources in the Big Data Era, Roles and Possibilities of Data Scientists,” *IT & Future Strategy*, Vol. 8, http://www.itsa.or.kr/bbs/board.php?bo_table=mmmmml&wr_id=357&page=2&ckattempt=1, 2012. (in Korean).
- [21] Matei, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A., *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute, 2011.
- [22] McClelland, D.C., “Testing for competence rather than for intelligence,” *American Psychologist*, Vol. 28, pp. 1-14, 1973.
- [23] McCormick, E.J., “Job and Task Analysis,” In M.D. Dunnette (Ed.), *Handbook of*

- Industrial and Organizational Psychology, Chicago: Rand McNally, pp. 651-696, 1976.
- [24] Metcalf, M. and Brenza, J., "Evaluating big data projects - success and failure using an integral lens," *Integral Leadership Review*, <http://integralleadershipreview.com/10945-evaluating-big-data-project-s-success-failure-using-integral-lens/>, 2013.
- [25] Park, C.W., Kim J.W., and Kwon, H.J., "An Empirical Research on Information Privacy Risks and Policy Model in the Big Data Era," *The Journal of Society for e-Business Studies*, Vol. 21, No. 1, pp. 131-145, 2016.
- [26] Park, J.S. and Han, S.G., "Job Analysis Guidelines for Vocational Education & Training Program Development," Korea Vocational Competency Development Institute, Working Paper, 2005.
- [27] Patil, D.J., *Building Data Science Teams*. O'Reilly, 2011.
- [28] Pereira, P., Cunha, J., and Fernandes, J.P., "On understanding data scientists," In *2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 1-5). IEEE, 2020.
- [29] Saif, N., Khan, M.S., Rehman, K., Rehman, S.U., Rehman, Z.U., Nawa, T., & Naqeeb, M., "Competency based Job Analysis," *International Journal of Academic Research in Accounting, Finance and Management Sciences*, Vol. 3, No. 1, pp. 105-111, 2013.
- [30] Sanchez, J.I. and Levine, E.L., "Is job analysis dead, misunderstood, or both?: New forms of work analysis and design," *Evolving Practices in Human Resource Management*, pp. 43-68, 1999.
- [31] Sanchez, J.I. and Levine, E.L., "What is (or should be) the difference between competency modeling and traditional job analysis?," *Human resource management review*, Vol. 19, No. 2, pp. 53-63, 2009.
- [32] Sanchez, J.I. and Levine, E.L., "The Rise and Fall of Job Analysis and the Future of Work Analysis," *Annual Review of Psychology*, Vol. 63, pp. 397-425, 2012.
- [33] Singh, P., "Job analysis for a changing workplace," *Human Resource Management Review*, Vol. 18, No. 2, pp. 87-99, 2008.
- [34] Sopa, A., Asbari, M., Purwanto, A., Santoso, P. B., Mustofa, D. H., Maesaroh, S., & Primahendra, R., "Hard skills versus soft skills: which are more important for Indonesian employee's innovation capability," *International Journal of Control and Automation*, Vol. 13, No. 2, pp. 156-175, 2020.
- [35] Sullivan, D., *Official Google Cloud Certified Professional Data Engineer Study Guide*. John Wiley & Sons, 2020.
- [36] Swan, A. and Brown, S., "The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs," *Technical Report*, University of Southampton, 2008.
- [37] Tammaro, A.M., Matusiak, K.K., Sposito,

- F.A., & Casarosa, V., "Data curators roles and responsibilities: an international perspective," *International Journal of Libraries & Information Services*, Vol. 69, No. 2, pp. 89-104, 2019.
- [38] Thompson, D.E. & Thompson, T.A., "Court standards for job analysis in test validation," *Personnel Psychology*, Vol. 35, No. 4, pp. 865 - 874, 1982.
- [39] Todd, P., McKeen, J.D., & Gallupe, R.B., "The evolution of IS skill: A content analysis of IS job advertisements from 1979 to 1990," *MIS Quarterly*, Vol. 19, No. 1, pp. 1-27, 1995.
- [40] Verma, A., Yurov, K.M., Lane, P.L., and Yurova, Y.V. (2019), "An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements," *Journal of Education for Business*, Vol. 94, No. 4, pp. 243-250, 2019.
- [41] Yoon, J. H., Ahn, J. Y. & Lee, C. C., "A review of the data scientist requirements: based on AHP technique," *Korean Review of Corporation Management*, Vol. 12, No. 1, pp. 41-62, 2021.

저 자 소 개



김향단
2014
2019
관심분야

(E-mail: hyangdan0129@gmail.com)
연변대학교 경영정보시스템 (학사)
한양대학교 경영정보시스템 (석사)
데이터분석, MIS



백승익
1990년
1993년
1998년
1998년~2000년
2000년~현재
관심분야

(E-Mail: sbaek@hanyang.ac.kr)
서강대학교 경영학과 (학사)
George Washington University (MBA)
George Washington University (Ph.D.)
Saint Joseph's University, 조교수
한양대학교 교수
Business Intelligence, Data Transformation, 서비스 혁신