

-Global Frontier Program-

국내외 데이터 활용 선진 사례 분석을 통한 효과적인 데이터 교육 방법



B.T.S (e-Business and IT management and Sociology)

팀장 이성범 201646006

팀원 류명기 201527014

이준혁 201646039

박예진 201946005

목차

I. 서론

1. 데이터의 중요성
2. 데이터 전문 인력 양성의 필요성

II. 본론

1. 대한민국 데이터 활용의 실태 및 인식 현황

- 1-1. 대한민국 데이터 활용 주요 분야
- 1-2. 데이터 활용 실태 및 인식
- 1-3. 데이터 활용에 대한 개인정보 인식

2. 대한민국 데이터 교육의 현황

- 2-1. 데이터 전문인력 양성 프로그램
- 2-2. 데이터 전문인력 자격 검증 프로그램
- 2-3. 효과적인 데이터 교육을 위해 나아가야 할 방향

3. 국내외 데이터 활용의 선진 사례

- 3-1. Kaggle
- 3-2. 국내 공공데이터 활용 사례
- 3-3. 해외 데이터 활용 사례

III. 결론

IV. 참고문헌

I. 서론

1. 데이터의 중요성

2020 년, 데이터는 세상을 바꾸고 있다. 우리의 모든 것은 데이터가 되고 기업과 정부는 이를 통해 새로운 정책을 개발하고 보다 나은 서비스를 제공한다. 경찰은 범죄 발생을 예측하고, 의료분야에서는 환자의 생명주기를 예측하면서 사고의 위험을 줄이고 있다. 코로나 발생 초기, 공공데이터를 활용한 마스크 맵은 전국적인 마스크 대란을 잠재우기도 했다. 이처럼 우리는 데이터를 활용해 사회의 불확실성을 감소시키고 더 나은 의사결정을 내릴 수 있게 되었다. 데이터는 단순히 한 개인의 편익을 증가시키고 삶의 질을 개선하는 것이 아닌, 앞으로 살아갈 시대의 삶의 방식과 방향을 완전히 변화시키고 불확실한 미래의 해결책으로 제시되고 있다.

코로나로 인해 세계는 위기에 봉착했지만 한편으로는 또 다른 기회를 맞이하고 있다. 비대면 문화가 발달함에 따라 온라인, 인공지능 산업이 급속하게 성장했고, 개인이 생산해내는 데이터 또한 방대하게 급증했다. 코로나 이후 불편하다고 생각했던 부분들은 이제 우리 삶에 어느 정도 정착했고, 우리가 불가능하다고 생각했던 기술들은 데이터를 통해 실현되고 구현되고 있다. 어쩌면 우리는 2020 년이 아닌 2025 년을 살아가고 있고, 세상은 이미 '포스트 코로나 시대'를 넘어서 '디지털 트랜스포메이션 시대'를 준비하기 위한 구체적인 논의를 진행하고 있다. 이러한 흐름 속에서 대한민국 또한 21 세기의 원유인 데이터를 통해 새로운 도약을 준비하고 있다. 정부는 데이터 3 법을 개정하고 디지털 뉴딜정책을 통한 미래의 비전을 제시하며 다양한 사업을 진행하고 있다.

2. 데이터 전문 인력 양성의 필요성

정부는 지난 17 년 12 월, 4 차산업 혁명 대응 계획을 발표하고 데이터 경제로의 전환을 선언했다. 문대통령은 직접 최고 수준의 AI 인재를 양성하겠다고 선언했으며, 지난해 12 월에는 인공지능 국가 전략도 발표했다. 이에 AI 대학원을 7 개 추가 지정했고, 이노베이션 아카데미를 개소하는 등 1000 여명의 빅데이터 및 AI 전문 인력을 양성하고 있다. 그러나 현장에서는 당장 필요한 인력도 심각하게 부족하다고 호소한다. 실제로 AI 와 같은 전문인력은 고사하고도 원재료인 데이터를 다룰 인재조차 부족하다고 전해지고 있다. 한국 데이터 산업 진흥원이 발표한 「2019 국내 데이터 산업 현황 조사」에 따르면 지난해 순수 데이터 직무 인력은 8 만 9000 여명 수준이며, 2020 년 기준으로 빅데이터 및 AI 산업에 부족한 인원은 약 4000 여 명이다. 향후 2023 년까지 2 만 2600 여 명이 더 부족할 것으로 추산된다. 또한 2018 년 과학기술정보통신부에서 발표한 「데이터 산업 현황조사」의 내용을 확인하더라도 <그림 1-1>과 같이 현재 빅데이터 수집·분석을 위한 시스템 도입, 원활한 시스템 운영을 위한 관심

증대에도 불구하고 빅데이터 시스템을 도입하지 않는 가장 주요한 이유는 “빅데이터 분석을 위한 전문인력의 부족”으로 나타났다.



<그림 1-1 빅데이터 미도입 이유(2018 년 데이터 산업 현황조사, 과학기술정보통신부)>

한국이 데이터 전문 인력 양성에 주춤거리는 동안 세계는 어땠을까? 2014 년부터 2018 년까지 전세계 데이터 전문인력의 증가 추이를 보면, 미국의 경우 1046 만명→ 1450 만 명으로 400 만명이 증가했고, EU 의 경우 582 만 명에서→723 만 명으로 140 만명이 증가했으며, 일본의 경우만 하더라도 334 만 명에서→411 만명으로 80 만명이 증가했다. 하지만 한국의 경우 26 만 7000 명에서→31 만 8000 명으로 약 5 만명 증가에 그쳤다. 실제 지난 해 12 월 국회입법조사처에서는 미국 등 경쟁국은 빅데이터 기술 수준을 기반으로 인공지능 발전을 추진 중이지만, 한국은 빅데이터 기술 수준이 미국 대비 83.4%가 낮다고 발표되기도 했다.

그 결과 스위스 국제경영개발대학원의 「2018 전 세계 디지털 경쟁력 순위 보고서」에서 발표된 데이터 활용 순위에서 대한민국은 전체 63 위중 31 위를 기록했다. 세계적인 기술력과 자본을 보유하고 있음에도 데이터 활용 순위가 낮은 이유는 미국, 영국과 같은 데이터 선진국에 비해 전문 인력이 턱없이 부족하고 이들을 양성할 수 있는 데이터 교육의 기반이 제대로 갖춰져 있지 않기 때문이다.

앞으로 세계의 모든 국가와 기업의 경쟁력은 데이터를 얼마나 잘 활용하는지에 따라 결정될 것이다. 물론 이 활용에 기반에는 데이터를 다룰 수 있는 전문인력이 확보 되어야하고, 이들을 양성할 수 있는 기본적인 교육이 잘 갖춰져 있어야 한다. 현재 대한민국은 미래 시대를 맞이할 본격적인 준비를 하고 있다. 디지털 트랜스포메이션 전환이라는 큰 과제 속에서 가장 중요한 것은 해당 영역의 전문 인력을 양성하는 것이다. 우리는 본 조사를 통해 미래 사회의 국가와 기업의 경쟁력인 데이터에 대해, 현재 대한민국에서 소비되는 데이터 활용과 실태, 인식, 교육의 현황을 알아보고 해외의 데이터 선진 활용 사례를 분석하여 데이터 인재를 보다 잘 양성하기 위한 개선된 교육 방안을 제시하려 한다. 나아가 한신대학교의 AI 빅데이터 센터 설립과 “창의융합 PLATFORM 교육혁신”이라는 교육 방향성과 비전 확립에 있어서 본 조사가 작게나마 도움이 되었으면 한다.

II. 본론

1. 대한민국 데이터 활용의 실태 및 인식 현황

대한민국 데이터 활용의 실태 및 인식 현황을 분석하고자 데이터가 활용되는 주요 분야 및 사례, 국내 기업 및 공공기관 데이터 활용 실태 연구, 그리고 국민들의 개인정보 활용에 대한 인식에 대하여 조사를 했다.

1-1. 대한민국 데이터 활용 주요 분야

대한민국의 데이터 활용 분야는 문화관광, 식품건강, 교육, 농축수산, 사회복지, 산업고용, 보건의료, 교통물류 등이 있다.

먼저 문화관광 분야에서는 전국 도서관 대출 현황을 통해 개인별 도서 추천 서비스를 제공하고 있다. 두 번째로 식품건강 분야에서는 화장품 특정성분 정보를 활용하여 화장품 성분 분석을 통해 개인 맞춤형 화장을 추천하고 있다. 세 번째로 교육 분야에서는 공공교육 데이터와 학교일정을 활용하여 각 학교마다 교육일정을 비교하고 시험정보를 제공하고 있다. 네 번째로 농축수산 분야에서는 도매시장 경락 가격정보를 통해 실시간 경매 정보 검색을 할 수 있고 거래현황 정보를 제공하고 있다. 다섯 번째로 사회복지 분야에서는 장애인 편의시설, 장애인 활동지원 기관 정보를 활용하여 휠체어 아동 전용 네비게이션과 자원봉사 활동 정보를 제공하고 있다. 여섯 번째로 산업고용분야에선 전자공시를 통해 취업준비생에게 객관적이고 신뢰성 높은 기업정보를 제공한다. 일곱 번째로 보건의료 분야에서는 전국병원 및 약국정보, 환자 데이터 정보를 활용하여 자신의 위치 기반 병원 혹은 약국 정보를 제공하고 상황별 환자 맞춤형 의료 정보 분석을 하고 있다. 마지막으로 교통물류 분야에서는 주차장 정보와 도로이점표 및 교통정보 데이터를 통해 인근 주차장 검색과 빠른 길 안내 서비스를 제공하고 있다.

현재 데이터의 활용 사례를 보면 수많은 정형, 반정형, 비정형 데이터들이 존재하지만 아직까지는 데이터 활용은 대부분 조회하거나 단순 분석으로만 사용되고 있다는 것을 알 수 있다. 따라서 데이터를 활용하여 창의적이고 독창적이며 고부가가치를 제공해주는 서비스를 만드는 것은 아직 부족하다고 할 수 있다.

1-2. 데이터 활용 실태 및 인식

그렇다면 우리나라에서 데이터는 얼마나 활용되고 있을까? 첫 번째로 국내 기업들을 살펴보고자 한다. NIA(한국 지능정보 사회 진흥원)의 조사에 따르면 국내 기업의 빅데이터 활용도는 10 퍼센트에 미치지 못하며, 종업원 수 100 명 이상의 기업은 4.3 프로에 그쳤다고

한다. NIA 가 진행한 빅데이터 시장현황조사에 따르면 가장 큰 이유로 지목된 것은 빅데이터라고 부를 만한 데이터가 없다는 것이었다. 분석 도입 효과가 확실하지 않다는 점과, 데이터 분석에 대한 무관심 등이 그 뒤를 이었다. 이를 통해 우리나라에서 데이터를 충분히 활용하기에는 빅데이터 자체에 대한 무관심과 데이터 부족, 그리고 이로 인하여 발생하는 데이터 분석에 대한 회의감이 문제라는 것을 알 수 있다.

두 번째로 A 교육청 소속 공무원들을 대상으로 온라인 설문조사를 통하여 진행된 연구를 분석했다. A 교육청 공무원들은 76.8% 이상이 자신의 업무를 위하여 데이터를 활용하고 있다. 그런데 자신의 업무를 할 때의 데이터 분석 능력이 어떠 한지를 묻는 문항에서는 응답자의 50% 이상이 '보통이다'를, 30%는 '전혀 그렇지 않다'와 '그렇지 않다'를 선택함으로써 자신 스스로 분석 능력이 낮다고 평가하였다. 이때, 데이터 활용에서 가장 어려운 부분이 무엇인지를 묻는 문항에서는 39.5%로 '수집 및 정제의 어려움'이 가장 높았고 19.5%로 '관련 전문지식부족'이 두 번째, 16.1%로 '필요한 데이터 소스 정보 부족'이 세 번째로 높았다. 빅데이터에 대한 교육청 공무원들의 이해 및 활용에 대한 인식을 묻는 문항에서는 빅데이터에 대한 이해 수준은 평균 3.03 점으로 보통 수준이었다. 하지만 빅데이터 활용 필요성과 관심도, 전문조직 필요성은 각각 4.06 점, 3.64 점, 3.66 점으로 빅데이터 이해 수준에 비해 상대적으로 높았다. 빅데이터 활용을 위하여 요구되는 사항이 무엇인지를 묻는 문항에서는 56.2%로 데이터 수집 및 관리 시스템 등 인프라 구축이 가장 필요하다고 말하고 있다. 두 번째로 32.5%로 '전담 조직 구축'과 32.3%로 '빅데이터 비전 및 목표, 방향 수립' 이 필요하다는 응답의 비율이 비슷하게 나왔다. 세 번째로 26.4%로 '빅데이터 활용을 위한 전문인력 확보'와 15.9%로 '데이터 수집'이 필요하다는 응답이었다.

이러한 결과를 통해 교육청 공무원들은 자신들이 업무에 활용하는 데이터의 유형과 출처는 다양하지만 데이터 수집 및 정제와 같은 지식과 기술이 부족하여 데이터 활용을 제대로 하지 못하고 있음을 알 수 있다. 또한 효율적인 데이터 활용을 위한 교육 인프라도 부족하여 많은 어려움을 겪고 있다. 이는 자신의 직무에서 데이터를 활용하고자 하는 관심과 필요성은 높지만 구성원들의 데이터에 대한 지식 수준과 기술 수준이 부족하여 효과적으로 데이터를 활용할 수 없다는 것으로 판단할 수 있다.

1-3. 데이터 활용에 대한 개인정보 인식

데이터 활용에 대한 국민들의 의식은 어떨까? 경기연구원에서 우리나라 국민 1000 명을 대상으로 한 개인정보 활용 관련 인식조사를 한 결과, 기술개발 및 경제발전을 위한 데이터 활용은 필요하다고 응답한 비율이 70%를 넘었으며 특히 개인정보 활용 목적이 공익적 목적에 기여할 경우, 정보를 제공할 의향이 있다는 응답이 84%에 달했다. 그 외에도 개인정보가 좋은 방향으로 활용될 시 데이터 제공 의향이 있다는 응답이 모든 항목에서 절반을 넘어섰다.

그러면 국민들이 개인 정보 제공 및 활용 시에 가장 중요하게 여기는 요소는 무엇일까? 설문에 따르면 개인정보 제공 및 활용 시에 고려하는 가장 중요한 요소는 '개인정보 활용 모니터링'이었으며, '개인정보보호 신뢰성과 안정성', '개인정보 활용 필요성과 목적', '개인정보 활용범위의 결정' 등 개인정보 보호와 연관되는 요소들이 그 뒤를 이었다. 또한, 적절한 인센티브가 제공될 시 개인정보를 제공할 의향이 있다는 응답 또한 절반 가까이 차지했다. 그 인센티브가 무엇 인지에 관한 설문 결과로는 56 퍼센트에 달하는, 절반 이상의 사람들이 '현금'이라고 응답하였으며, 그 후로는 '맞춤형 서비스', '세금 감면', '상품권' 등이 뒤를 이었다.

데이터를 효과적으로 활용하기 위해서 개인정보 활용에 대한 인식에 대한 개선이 필수적으로 이루어져야 한다. 위 설문 조사 결과를 바탕으로 봤을 때 개인정보 활용에 대한 인식이 개선되기 위해서는 공익의 목적으로 사용되고 이에 대한 홍보가 제대로 이루어지고 개인 정보 활용 범위를 명확하게 정하고 공지하며, 해당 정보가 얼마나 안전하게 보호되고 있고 어디에서 활용되고 있는 지를 이용자가 직접 확인할 수 있는 시스템 구축이 무엇보다 중요하다. 그리고 데이터 활용에 대한 인센티브가 마련되어야 하며 대부분이 현금을 원하지만 이는 너무나 큰 재정이 뒷받침되어야 하기 때문에 맞춤형 서비스 제공, 세금 감면 등을 통해서 많은 사람의 요구를 만족할 수 있는 방안을 찾는 것이 중요하다. 이런 제도적 뒷받침이 완료된다며 효과적인 데이터 활용을 위한 개인정보 제공 및 활용의 인식이 개선이 될 것이다.

우리나라에서도 수많은 데이터가 존재하며 활용되고 있으나 실제로 데이터를 활용하는 교육청에서의 연구를 통해 보았듯이 아직 그 활용성이 충분하지 않음을 알 수 있다. 하지만 국민들은 데이터가 안전하게, 공익을 위하여 사용된다면 개인정보를 제공함에 있어 긍정적이기 때문에 이에 대한 시스템을 구축함으로써 데이터에 대한 인식뿐만 아니라 양과 질을 개선시키면 자연스럽게 데이터에 대한 관심과 교육 확대에 이어질 것으로 기대된다.

2. 대한민국 데이터 교육의 현황

대한민국 데이터 교육의 현황을 분석하고자 첫번째로 정부, 대학, 민간에서 주도하는 전문 인력 양성 프로그램에 대해서 조사를 진행했다. 두번째로 데이터 분석 전문가에 대한 자격 검증 프로그램에 대해서 조사를 진행했다. 마지막으로 조사를 바탕으로 효과적인 데이터 교육을 위해 나아가야 할 방향에 대해서 제시했다.

2-1. 데이터 전문인력 양성 프로그램

데이터 전문인력 양성 프로그램은 크게 정부, 대학, 민간 주도의 프로그램으로 나뉘어 진다. 최근에는 정부에서 제공하는 국민내일배움카드를 통해서 민간에서 운영하는 데이터 전문인력 양성 프로그램을 정부의 지원금을 바탕으로 최대 전액까지 지원받아 무료로 교육을 받을 수 있다. 이러한 측면에서 볼 때 정부 주도와 민간 주도의 프로그램에서 교육비 부담에 대한

차이는 사실상 없다고 볼 수 있으며 큰 틀에서는 정부의 주도로 민간 프로그램을 운영한다고 볼 수 있다. 또한 정부 주도의 프로그램이 대학과의 협약을 통해서 대학에서 주도로 하여 운영하는 경우도 존재한다. 따라서 정부, 대학, 민간 주도의 프로그램들은 큰 틀에서는 정부에서 운영을 하지만 운영 자체는 대학, 민간 등의 각 기관에서 맡아서 독자적으로 운영하는 것으로 볼 수 있다. 따라서 커리큘럼 및 운영 측면에서 차이를 보인다. 이에 정부, 대학, 민간 주도의 프로그램을 나누어 각 프로그램의 특징에 대하여 분석을 진행하였다.

첫번째로 정부에서 주도하는 데이터 전문인력 양성 프로그램이다. 정부에서는 주로 한국 데이터산업진흥원에서 주관하여 데이터 전문 인력 양성 프로그램을 운영하는 중이다. 운영하는 프로그램으로는 크게 빅데이터 아카데미와 빅데이터 청년인재 일자리 연계 과정이 있다.

그 중 빅데이터 아카데미는 시장수요에 대응한 인력을 양성하고, 전문 직업으로서의 육성을 지원하기 위하여 실무 경력을 3 년 이상 보유한 산업계 재직자를 대상으로 운영되는 프로그램이다. 2020 년 5 월 기준 75 회 운영되었으며 1792 명의 수료생을 배출한 프로그램이다. 프로그램의 교육 과정은 크게 직무별 전문인력 양성 프로그램과 산업별 융합인력 양성 프로그램으로 나뉘어진다. 직무별 전문인력 양성 프로그램은 세부적으로 빅데이터 기획, 기술, 분석 전문가와 데이터 거래사 양성으로 나뉘어지며 산업별 융합인력 양성 프로그램은 세부적으로 금융, 유통, 제조, 의료 빅데이터 전문가로 나뉘어진다. 각 과정들은 크게 오프라인 집체교육을 거쳐서 교육받은 내용을 바탕으로 프로젝트를 만들고 그 프로젝트를 발표하는 식으로 총 5 주간 운영된다. 각 과정마다 세부적으로 배우는 내용은 달라지지만 크게 프로그래밍 언어의 경우 Python 과 R 을 학습하고 세부적인 학습 내용은 머신러닝, 딥러닝, 통계학, 데이터 활용 사례에 대하여 학습을 진행한다. 총 5 주간의 교육 중 프로젝트 발표와 계획을 제외하고는 실제로 교육을 받는 기간은 2주에 불과한 프로그램이다. 재직자를 대상으로 하는 수업이지만 교육의 취지는 데이터 전문가 양성이다. 하지만 전문가를 양성하는 것에 5 주의 교육 기간은 턱없이 부족하며 각 학습 내용도 2 주 동안 배우기에는 시간 상으로 매우 부족하다고 생각된다. 따라서 현 프로그램의 경우 전문성과 실무 경험을 중시하는 현재의 기업 채용 문화에 초점은 맞춰져 있지만 그 기대에 부응하기에는 시간 상으로 매우 턱없이 부족한 프로그램이라고 생각된다.

두번째로 대학에서 주도하는 데이터 전문인력 양성 프로그램이다. 대학은 데이터 관련 학과를 개설하여 교육을 하는데 중점을 두고 있으며 크게 학부 과정과 대학원 과정으로 나뉘어진다.

학부과정에서는 기존의 데이터마이닝과 정보통계처리학 등의 교육 영역을 확장하여 빅데이터와 다른 학문을 융합한 경영정보, 마케팅 등 다양한 교육커리큘럼을 개발하고 시행하고 있다. 학부에서는 별도의 빅데이터 전공학과를 개설하기보다는 연계 전공을 통해 관련학과 강의를 제공함으로써 데이터 전문인력이 되기 위해 필요한 기초역량과 창의적 사고를 위한 기반을 마련에 중점을 두고 있다. 또한 전문성과 실무 경험을 중시하는 현재

기업의 채용 문화에 대응하기 위하여 교과 과정에 산업에서 요구하는 전문인력 역량에 맞춰 융합형 실무형 인재육성을 지향하며 기업체 실무 전문가가 참여하여 수업을 진행하는 방식으로 운영되고 있다. 하지만 대학에서 다루는 기술과 실제 실무에서 적용되는 기술 간에는 언제나 큰 차이가 존재한다. 또한 학부를 졸업하고 바로 취업을 할 수가 없어서 경력을 쌓기 위해서 대학원으로 진학을 결정하는 학생들도 많다. 이러한 측면에서 바라볼 때 현재 학부 수준의 교육이 크게 취업에 도움이 된다고 말을 할 수는 없다. 따라서 대학은 이러한 문제점을 인식하여 더욱더 취업과 실무를 중점으로 다루는 교육을 다룰 필요가 있다고 생각된다. 대학 교육과 실무와의 괴리를 좁히기 위해서는 대학 커리큘럼 개발에 기업과 협력을 할 필요성이 높다고 생각되며 기업과의 활발한 인턴쉽 및 협력을 통해서 학생 스스로가 전문성의 필요성을 인식하며 스스로 공부의 방향을 결정할 수 있도록 도와주도록 하는 교육이 필요하다고 생각된다.

마지막으로 민간에서 주도하는 데이터 전문인력 양성 프로그램이다. 민간 주도의 프로그램은 국내외 IT 기업 또는 민간 교육전문기관에 의해 운영되고 있다. 교육은 재직자의 실무 역량 강화와 취업준비생들을 위한 역량 향상에 목적을 두고 진행되고 있다. 민간 주도의 프로그램의 경우 다른 프로그램보다 가장 전문적인 교육을 진행하고 있으며 기업과의 연계를 통해서 취업 성공률과 만족도 또한 높은 편이다. 하지만 민간 주도의 프로그램의 경우 높은 교육비가 들어간다는 단점과 실무에 초점을 맞춘 나머지 기술력은 갖췄지만 이론에 대한 전문성은 부족하다는 단점이 존재한다. 민간의 교육전문기관인 패스트캠퍼스의 경우 데이터 사이언스 직무 교육을 통해 데이터 분석, 머신러닝, 딥러닝, 텍스트 마이닝 등의 강좌를 제공하며 직접적인 교육성고가 제시될 수 있는 실무 분야로 한정하여 프로그램이 운영되며 강사로는 산업 현장의 평판과 인지도를 고려하여 우수한 기술력을 갖춘 강사를 섭외하고 교육 수료 후 6개월 동안 취업 도움을 같이 제공한다.

2-2. 데이터 전문인력 자격 검정 프로그램

국가에서 인정하는 데이터 전문인력 자격 검정 프로그램으로는 한국데이터산업진흥원에서 주관하는 자격증이 존재한다. 한국데이터산업진흥원은 산업 수요에 부응하는 인재 양성을 위해 기업의 데이터에 관한 모든 구조를 체계화 하여 설계하는 능력을 검정하는 데이터아키텍처 자격시험과 데이터베이스 개발의 필수 언어인 SQL 활용 능력을 검정하는 SQL 자격시험과 과학적 의사결정을 지원하기 위해 데이터를 활용하여 분석하는 역량을 검정하는 데이터 분석 자격시험을 제공하고 있다.

데이터아키텍처 자격시험은 DAsP 와 그 상위 단계인 DAP 로 나뉘어진다. 다루는 과목은 전사아키텍처 이해, 데이터 요건 분석, 데이터 표준화, 데이터 모델링, 데이터베이스 설계와 이용, 데이터 품질 관리 이해로 총 6 과목이다.

SQL 자격시험은 SQLD 와 그 상위 단계인 SQLP 로 나뉘어진다. 다루는 과목은 데이터 모델링의 이해, SQL 기본 및 활용, SQL 고급활용 및 튜닝 실무로 총 3 과목이다.

데이터 분석 자격시험은 ADsP 와 그 상위 단계인 ADP 로 나뉘어진다. 다루는 과목은 데이터 이해, 데이터 처리 기술 이해, 데이터 분석 기획, 데이터 분석, 데이터 시각화로 총 5 과목이다.

각 자격증의 응시자격은 하위 단계의 경우 제한이 없지만 상위 단계의 경우 하위 단계의 자격증을 필수로 취득해야 하며 박사, 석사 및 실무 경력을 요구한다는 측면에서 매우 전문적이며 체계적인 자격증이다. 또한 대한민국 최초이며 유일한 국가공인 데이터 전문인력 검증 자격증이라는 측면에서 매우 메리트가 높다.

2-3. 효과적인 데이터 교육을 위해 나아가야할 방향

현재 데이터 산업의 발전과 그 중요성이 나날이 높아져 가는 상황에서 기업과 국가에서는 전문성을 갖춘 데이터 전문 인력을 필요로 하고 있으며 이에 대한 수요도 폭발적으로 증가하고 있다. 따라서 효과적인 데이터 교육을 통하여 국가 발전에 이바지할 수 있는 데이터 전문 인력 양성의 필요성이 매우 높아졌다. 현재 우리는 이러한 수요에 빠르게 대응하기 위해서 데이터 전문 인력 양성을 위한 다양한 교육, 검정 프로그램을 제시하고 있다. 하지만 각 기관마다 주도하는 프로그램마다 장단점이 존재하며 이론과 기술을 갖춘 완벽한 데이터 전문 인력 양성에는 부족함이 있다고 생각된다.

정부의 프로그램의 경우 재직자를 대상으로 진행하며 프로젝트 위주의 실무 중점에 초점을 두고 있지만 교육의 기간이 매우 짧아 프로젝트와 교육이라는 두 가지 목적을 모두 이루기에는 어려운 것으로 판단된다.

대학의 프로그램의 경우 실무 능력을 기르기 위해서 커리큘럼을 준비는 하지만 대학이라는 한계가 존재하기 때문에 실무 능력을 기르기 것은 현실적으로 매우 어려우며 그 한계가 존재한다. 따라서 현재 연구에 종사하는 교수들을 중심으로 이론 분야에 대한 학습과 다양한 전공과 융합하여 그 분야에 대한 기반 지식을 쌓는데 초점을 맞추고 있다.

민간의 프로그램의 경우 실무에 종사하는 강사를 영입하여 실무 능력 향상에 초점을 맞추고 교육 후에 취업을 관리해주는 시스템을 구축하여 이론보다는 실무와 현장 중심의 교육에 중점을 두고 있다.

진정한 전문성을 갖춘 데이터 전문인력은 이론과 실무를 섭렵하며 스스로 프로젝트를 계획하고 마무리할 수 있는 능력을 갖춘 사람이라고 생각된다. 따라서 기업과 국가에서 원하는 진정한 전문성을 갖춘 인력을 양성하기 위해서는 정부, 민간, 대학이 협력하여 각 프로그램의 장점을 바탕으로 하나의 프로그램으로 합치며 기간을 길게 설정하여 전문인력을 양성하는 것이 좋다고 생각된다.

예를 들어 1 년 프로그램의 경우 대학생은 학교를 다니는 학기 중에는 대학에서 연구에 종사하는 교수님들의 강의 바탕으로 이론에 대한 교육을 받고 학기가 끝난 방학 기간에는 민간에서 진행하는 프로그램에 참여하여 실무 중심의 교육을 받아 실무 능력을 기르고 1 년 간의 교육 과정이 끝나면 배운 내용을 바탕으로 정부에서 주도하는 프로그램에 참여해 정부에서 제시하는 문제 해결이 필요한 프로젝트를 직접 계획하고 마무리하는 과정으로 진행을 한다. 그 후 자격 검정을 통해서 자격증까지 취득한다면 국가와 기업이 생각하는 진정한 전문성을 갖춘 인재로 거듭날 수 있다고 생각된다. 전체적인 프로그램의 관리를 정부가 주도한다면 참여 기관과 기업 등의 섭외는 어렵지 않다고 생각된다. 현재처럼 정부, 대학, 민간 기관이 독자적으로 운영하는 것이 아닌 다같이 협력하여 하나의 프로그램으로 만들어 실행한다면 데이터 경제 시대에 전문성을 갖춘 데이터 전문 인력을 양성할 수 있을 것이다.

3. 국내외 데이터 활용의 선진 사례

현재 국내외로 데이터를 활용한 사례가 점차 증가하고 있고, 그 분야 또한 다양하다. 데이터의 제공을 스스로 결정하여 데이터를 활용 유무를 선택하는 '마이 데이터'가 증가하는 추세이며, 많은 사람들이 자유롭게 사용할 수 있는 공공 데이터를 활용하는 것 또한 증가하고 있다. 데이터를 다양한 분야에서 어떤 방식으로 사용하며 새로운 가치 창출 및 환경의 개선을 불러왔는지 사례를 통해 분석해보려 한다.

3-1. Kaggle

Kaggle 은 2010 년에 설립된 데이터 분석 및 머신러닝에 대한 학습 플랫폼이자 경쟁을 할 수 있는 대회 플랫폼이다. 기업 및 단체에서 데이터와 해결과제를 등록하면, 데이터 분석에 관심이 있는 누구나 문제에 대한 해답을 제출할 수 있다,

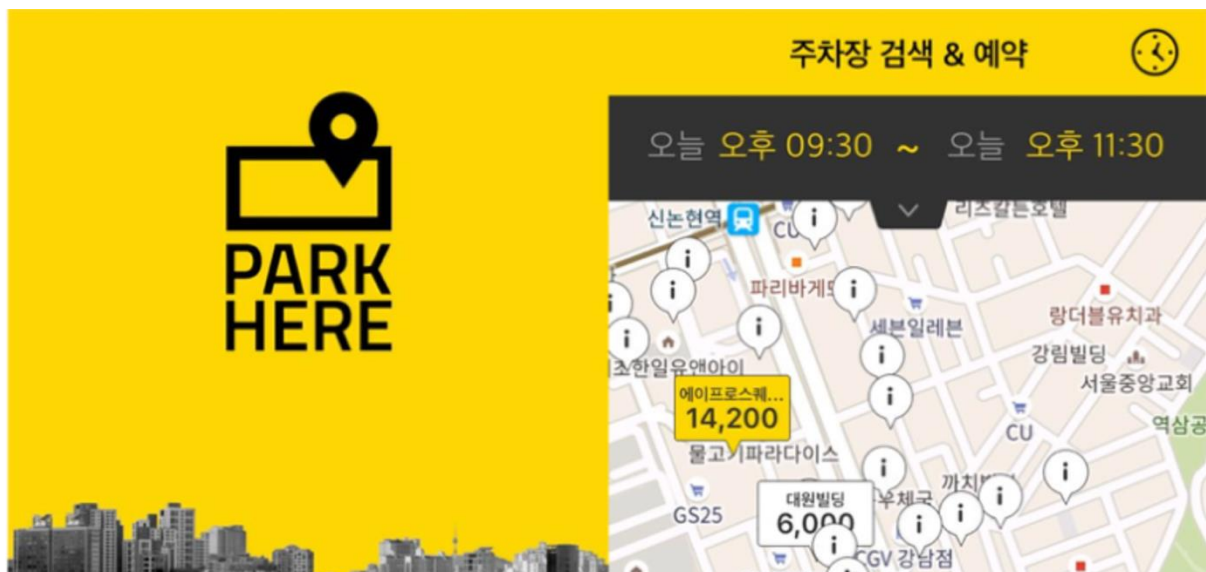
지금도 데이터는 생성되고 있으면 종래의 방법으로는 수집, 저장, 가공, 분석이 어려운 방대한 규모의 데이터, 즉 빅데이터가 만들어지고 있다. 빅데이터를 제대로 분석한다면 정말 많은 문제들을 해결할 수 있다. 이에 데이터를 분석하고 예측 모델을 도출해 낼 수 있는 전문가의 가치는 매우 높아지고 있다. 하지만 기업의 경우 이러한 분석을 진행할 전문가가 부족하거나 없는 경우가 매우 많다. 이러한 순간에 필요한 것이 바로 Kaggle 이다. 기업이 Kaggle 에 빅데이터를 제공해주면 캐글은 이를 온라인에 공개하여 세계 각지의 수 많은 데이터 과학자들로 하여금 문제를 해결할 수 있도록 해준다. 캐글은 한마디로 세계 최대의 데이터 과학자의 커뮤니티이다.

캐글을 통해서 기업은 자신이 해결할 수 없는 문제를 집단지성의 힘을 빌려서 해결할 수 있으며 캐글을 이용하는 개인들은 다른 사람들의 분석 내용을 통해서 학습하고 스스로가 프로젝트를 구성하여 문제를 해결함으로써 데이터 분석에 대한 경험과 실력을 모두 쌓을 수

있다. 따라서 캐글을 통해서 기업과 개인은 모두 자신의 역량을 향상시킬 수 있으며 데이터를 효율적으로 활용할 수 있다.

3-2. 국내 공공데이터 활용 사례

공공데이터를 활용한 대표적인 사례는 바로 화장품 비교 모바일 앱인 “화해”이다. 화해는 화장품 성분을 분석하여 소비자들에게 알기 쉽게 제공하는 모바일 앱으로서 네이버 트렌드리포터가 믿고 쓰는 앱으로 선정될 만큼, 유익한 정보를 제공하면서 많은 이용자를 보유하고 있다. 화해는 식약처의 화장품 정보와 특성 정보, 화장품 전 성분 정보 이 3 가지의 공공데이터를 기반으로 개발 되었다. 화해를 개발한 버드 앱 이웅 대표는 “엄밀히 말해 공공데이터가 없었다면 화해 앱은 탄생할 수 없었다”면서 공공 데이터에 대한 공을 높였다. 이처럼 공공데이터는 지금까지는 경험해보지 못한 분야에서 새로운 비즈니스 모델의 구축하는데 활용되고 있다. 또한 주차 문제를 해결하기 위한 새로운 솔루션으로 각광받고 있는 모바일 앱 “파크히어” 또한 공공데이터를 활용하여 새로운 비즈니스 모델을 구축한 좋은 사례로 손꼽힌다.



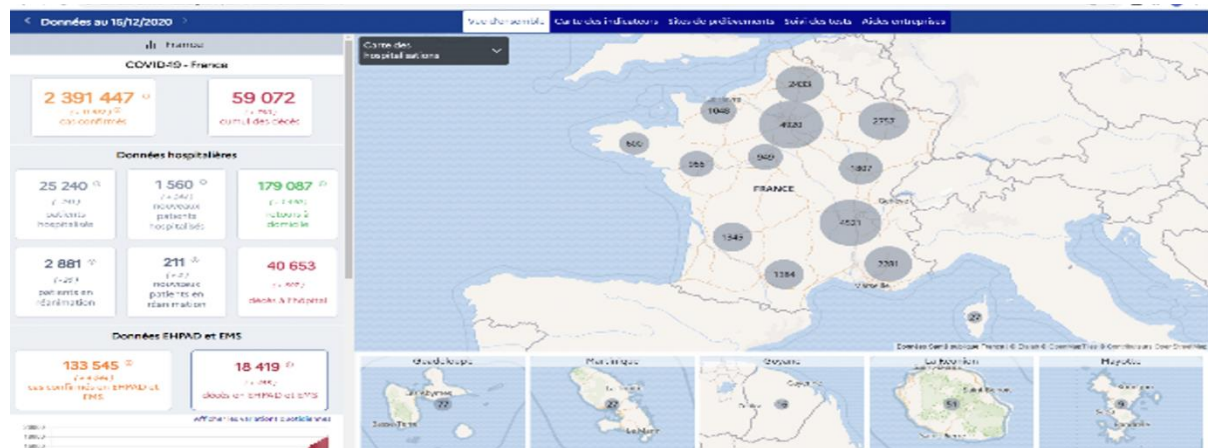
<그림 3-2 공공데이터를 활용하여 새로운 비즈니스 모델을 창출한 “파크 히어”>

‘파크히어’는 운전자들에게 목적지 주변의 빈 주차공간을 알려주고 주차장 할인 예약, 결제, 발권까지 한 번에 해결할 수 있는 모바일 앱이다. 빈 주차공간을 이용자들에게 제공하면서 수익을 창출할 수 있어 주차장 사업자들과 이용자들이 효율적으로 주차공간을 찾을 수 있게 도와주는 고마운 앱으로 평가 받고 있다. ‘파크히어’는 서울시의 구영, 시영주차장 정보와 국가교통데이터베이스의 전국 통계 주차장 공공데이터를 기반으로 개발되었다. 이러한 ‘파크히어’는 출시 일주일 만에 1,700 건의 다운로드를 돌파하면서 이목을 끌었고, 현재도 많은 이용자들이 사용하고 있다. 이처럼 점차 공공데이터를 활용하면서 새로운 가치를 창출하는

사례가 다양해지고 있다. 이를 기반으로 개발자는 수익을 창출하고, 이용자들은 보다 편리한 서비스를 이용할 수 있게 되면서 공공데이터는 새로운 블루오션으로 각광받고 있다.

3-3. 해외 데이터 활용 사례

① 프랑스의 공공 데이터 활용사례



<그림 3-1 공공데이터를 활용한 프랑스의 코로나 맵>

프랑스는 코로나 19의 발생 이후 관련 데이터들을 시민사회와 조직, 개발자 등의 민간 부문에서도 재사용 가능한 형식으로 제공했다. 현재까지도 코로나 19 환자 데이터 외에도 병원 응급실 데이터 등을 즉시 활용 가능한 CSV 파일로 개방 중이며, 그 외 다양한 자료들은 HTML, XLS, TXT, PDF 등으로 제공하며 지자체, 기업, 시민들이 목적에 따라 데이터를 자유롭게 사용하면서 새로운 가치를 창출 하고 있고, 실제 이를 이용하여 다양한 활용 사례가 나타나고 있다.

② 미국의 공공 데이터 활용사례

해외 데이터 활용 사례로는 미국의 스마트공시(Smart Disclosure)가 있다. 스마트 공시는 소비자가 의사결정과 제품 및 서비스 이용시 다양한 데이터를 활용할 수 있도록 소비자에게 데이터 접근 기회를 확대해 주는 것을 말한다. 정부는 공공기관이 보유한 정보를 정보주체에게 제공하고 민간과 협력을 진행하며, 민간 기업은 정보를 정보주체에게 제공하고 활용을 통해 소비자들에게 맞춤형 서비스를 제공한다. 스마트 공시는 Blue Button(건강 데이터)과 Green Button(에너지 데이터)이 등이 있다. Blue Button 은 개인정보 접근을 허용하는 기관(공공 또는 민간) 웹 사이트에서 '파란 버튼'을 클릭하여 개인 건강 데이터를 다운로드 할 수 있는 권한을 부여한다. 이를 통해 개인은 건강 데이터를 다운로드 받고, 기업이 제공하는 데이터 관련 서비스를 활용하거나 공유할 수 있다. Green Button 은 웹사이트의 '녹색 버튼'을 클릭하여 에너지 사용 관련 데이터에 접근할 수 있는 것으로 에너지 공급자 및 소비자는 실시간으로

에너지 사용 현황을 확인할 수 있으며, 소비자는 자신의 에너지 사용 데이터를 다운로드 및 공유할 수 있다.

Ⅲ. 결론

2019 년 국내 빅데이터 및 분석 시장은 전년 대비 10.9% 증가한 1 조 6744 억원을 기록했다. 국내 빅데이터 시장은 2023 년까지 연평균 성장률 11.2%를 기록하며 2 조 5692 억원의 규모에 달할 것을 전망된다. 국내 빅데이터 시장은 미래가 유망한 분야로서 손꼽히지만 그럼에도 시장의 수요를 채울 전문인력의 부재는 여전히 해결해야 할 중요한 문제로 남아있다. 국민의 95%가 스마트폰을 이용하는 IT 강국인 우리나라는 지금도 수많은 데이터가 생성되고 있다. 그리고 데이터를 활용하기 위한 기본적인 제도와 틀은 구축이 어느정도 구축이 완료되었다. 하지만 현재 이러한 데이터를 바탕으로 고부가가치를 창출할 수 있는 데이터 전문 인력은 부족한 실정이다. 따라서 데이터 활용 능력이 곧 국가의 경쟁력인 시대에 데이터 전문 인력을 양성하기 위한 효과적인 교육이 필요하다.

먼저 가장 중요한 것은, 전문인력을 양성할 교육 시스템의 개선과 확장이다. 현재는 데이터를 활용하기 위한 기본적인 제도와 기술력, 자금은 준비되어 있지만 빅데이터 전문 인력이 턱없이 부족한 것이 가장 큰 문제점이다. 대한민국의 경우 현재 정부, 대학, 민간기관의 주도로 하여 데이터 전문 인력 양성 프로그램을 진행하고 있다. 각 프로그램마다 장단점이 존재하며 독립적으로 운영 되고 있다. 결국 진정한 전문성을 갖춘 데이터 전문인력은 이론과 실무를 섭렵하며 스스로 프로젝트를 계획하고 마무리할 수 있는 능력을 갖춰야 한다. 이러한 데이터 전문 인력은 양성하기 위해서는 정부가 주도하여 민간 프로그램의 장점과 대학 프로그램의 장점을 하나로 묶어 하나의 프로그램을 만드는 것이 필요해 보인다. 학기 중에는 이론을 중심으로 수업을 진행하고 방학 기간에는 실무 중심의 체험을 진행하면서 이론과 실무를 두루 섭렵하는 것이다. 이후 전체적인 프로그램이 종료되면 지난 과정들을 토대로 정부가 제시하는 문제를 배운 내용을 바탕으로 해결하는 프로젝트를 진행함으로써 국가와 기업이 원하는 데이터 전문 인력으로써 성장할 수 있을 것이다.

현재 고등학교 교과 과정도 다가오는 데이터 시대에 부응하고자 새롭게 변화하고 있는 중이다. 서울시 교육청은 2024 년까지 총 250 억원을 투입해 인공지능과 빅데이터 분야 학교를 10 개교 설립을 예정하고 있다. 2021 학년도부터는 모든 특성화고 신입생에게 인공지능 관련 과목을 필수 이수하도록 하고, 교원 대상 중장기 연수를 운영할 계획을 통해 미래 인재들에 대한 조기 교육을 계획하고 있다. 비록 특성화고등학교에서만 진행되고 있는 교육 과정이지만 앞으로는 더욱 확대될 것으로 예상된다.

하지만 무엇보다 가장 중요한 것은 데이터를 우리 생활 속에 접목시켜 다양한 사고를 할 수 있는 좋은 환경을 조성해주는 것이 필요하다. 특수한 학과를 졸업해야만, 관련 기관에서

근무를 해야지만 데이터를 접할 수 있는 것이 아니라 누구나 기본적으로 데이터를 쉽게 접할 수 있는 환경이 필요하다. 이러한 관점에서 데이터 경진대회 플랫폼인 “캐글”과 “공공데이터포털”은 가장 좋은 사례라고 생각된다. 학력, 성별, 전공 등에 국한없이 모두가 동등하게 참여할 수 있는 캐글의 데이터 경진 대회는 개인이 접해볼 수 없는 기업의 방대한 데이터들을 다뤄보면서 데이터에 대한 이해와 사고의 방향을 넓힐 수 있다. 또한 전세계의 데이터 분석가들의 정보 공유를 통해 현업에서 적용되는 실질적이고 효율적인 정보를 얻을 수 있다. 또한 공공데이터포털의 경우, 단순한 CSV 나 엑셀과 같은 데이터가 아닌 실시간으로 제공되는 오픈 API 데이터를 누구나 조회하고 이를 활용하면서 기존보다 더 많은 가치를 창출할 수 있다. 기업과 정부는 데이터를 공개를 하여 원하는 가치를 보다 다양한 관점에서 접근하여 문제를 해결할 수 있고 개인은 데이터 전문 인력으로 성장할 수 있는 발판을 마련하여 서로 성장하는 상생의 관계가 형성될 것이다.

데이터 전문 인력을 양성하기 위한 정부 주도하에 정부, 대학, 민간이 통합한 프로그램을 만들고 누구나 쉽게 데이터를 접하고 다룰 수 있는 환경이 마련된다면 대한민국이 앞으로 데이터를 다루는 능력이 곧 국가의 경쟁력이 되는 시대에 데이터 활용 선도국가로써 발돋움할 수 있을 것이다.

IV. 참고문헌

1. 신종호(2020), 교육청 공무원의 데이터 활용실태 및 교육 빅데이터 활용에 관한 인식 연구, 한국디지털정책학회
2. 정승화/도재우(2019), 빅데이터 분석가 양성과정 운영 사례 연구, 교육연구소
3. 한국데이터진흥원(2016), 데이터 전문인력 양성 방안 연구 결과 보고서
4. IDC(2020), 국내 빅데이터 분석 2019-2023 시장 전망 연구보고서
5. 송영아(2020), 데이터의 효율적 활용을 위한 빅데이터 분석 교육과정 사례 연구, 실천공학교육논문지
6. 위키백과, Kaggle
7. 슈어소프트테크, Kaggle 이 궁금해? Kaggle 의 모든 것!, 네이버 블로그
8. 정책공감, 공공데이터 개방이 만든 일상 혹시 내가 쓰는 앱도 공공데이터?, 네이버 블로그
9. 공공데이터포털, 공공정보로 탄생한 화해 앱 화장품 성분을 쉽게 볼 수 있죠!, 네이버 블로그
10. 한국데이터산업진흥원, <https://www.kdata.or.kr/>
11. 김윤미, 해외 마이데이터 사례 분석 및 국내 적용을 위한 시사점 도출, 한국신용정보원
12. 한국정보화진흥원 공공데이터본부 공공데이터 기획팀, 해외 코로나 19 관련 데이터 개방 현황 및 주요 질병관리 통계 기관의 데이터 개방 현황 제 20 호, NIA 한국정보화진흥원
13. [정책위키] 한눈에 보는 정책, 데이터경제, 대한민국 정책브리핑
14. 최재경, 빅데이터 분석의 국내외 활용 현황과 시사점
15. 이동철, 공공데이터 활용 및 사례, 한국도시행정학회