

빅데이터 양성 교육 교과과정 개선을 위한 회귀분석 기반의 만족도 조사에 관한 연구

A Study on Satisfaction Survey Based on Regression
Analysis to Improve Curriculum for Big Data Education

최현*

Hyun Choi*

〈Abstract〉

Big data is structured and unstructured data that is so difficult to collect, store, and so on due to the huge amount of data. Many institutions, including universities, are building student convergence systems to foster talents for data science and AI convergence, but there is an absolute lack of research on what kind of education is needed and what kind of education is required for students. Therefore, in this paper, after conducting the correlation analysis based on the questionnaire on basic surveys and courses to improve the curriculum by grasping the satisfaction and demands of the participants in the “2019 Big Data Youth Talent Training Course” held at K University, Regression analysis was performed. As a result of the study, the higher the satisfaction level, the satisfaction with class or job connection, and the self-development, the more positive the evaluation of program efficiency.

Keywords : Big Data, Training Course, Regression Analysis, Basic Surveys

* 정회원, 주, 경남대학교 공과대학 토목공학과 교수
E-mail: hhcoi@kyungnam.ac.kr

* Dept. of Civil Engineering, Kyungnam University.

1. 서론

빅데이터란 데이터의 양이 너무 방대하여 기존 처리방법이나 도구로 수집·저장 등이 어려운 정형 및 비정형 데이터를 말한다. 빅데이터는 크기(Volume), 속도(Velocity), 다양성(Variety)의 특징이 있다. 크기는 수십 테라바이트 혹은 수십 페타바이트 이상 규모의 데이터 속성을 가지는 특징이 있다. 속도는 대용량의 데이터를 빠르게 처리하고 분석할 수 있다. 그리고 통신기술의 발전으로 인해 국민과 기업의 각종 활동에 대한 방대한 빅데이터가 실시간으로 축적되고 있어, 세계 각국은 빅데이터를 활용한 신산업 창출과 공공정책 개발 등에 박차를 가하고 있다. 그러므로 AI, 디지털 사이언스 등 융복합 환경에서의 디지털 데이터는 매우 빠른 속도로 생산되기 때문에 실시간으로 저장, 유통, 수집, 분석처리가 가능한 성능을 갖추어야 한다. 다양성(Variety)은 복잡다묘인의 환경에서 생성되는 데이터를 의미하며 정형화의 종류에 따라 정형, 반정형, 비정형 데이터로 구분할 수 있다. 특히 넷플릭스와 아마존이 고객의 구매 패턴을 분석한 뒤 맞춤형 추천 서비스로 성공을 하자 글로벌 기업들이 데이터 사이언스에 주목을 하고 있다. 지금은 전 세계적으로 빅데이터의 실체에 대한 논란에서 벗어나 데이터의 실질적인 가치를 발굴하고 비즈니스로 연결시켜 데이터와 AI는 미래 산업의 핵심으로 성장하고 있다. 최근 기후변화와 맞물려 국토연구원에서는 빅데이터를 활용한 종합적 재난안전 대책 마련의 필요성에 대하여 언급이 되었다[1]. 또한 텍스트 마이닝을 활용한 우리나라의 기후변화 리스크 평가 및 적응 대책에 관한 연구에서 기존 연구들의 기후변화 리스크 평가에서 실제 관측 영향에 대한 결과 반영이 미흡했다는 문제의식에서 출발하여 우리나라 기후변화 관측영향(24년간 우리나라

기후변화 영향 관련 신문기사)의 리스크 결과값을 대상으로 주요 리스크를 도출하고, 유사 선행 연구와 비교·분석을 통하여 정책적 시사점을 제시하였다[2]. 최근 복잡해지는 현대사회의 뉴스 키워드를 시간적 흐름에 따른 빈도수와 상관관계로 분석을 위해 우리나라의 사회적 이슈 도출을 위한 뉴스 빅데이터 분석에 관한 연구에서는 5년 기준으로 이슈의 변화 흐름을 연구를 수행하여 변화 분석을 실시하여 과거와는 다른 새로운 갈등으로 사회문제가 심화될 수 있는 구조로 변화됨을 밝혀내었다.[3]

그리고 빅데이터 분석가 양성과정 운영사례에 관한 연구에서는 빅데이터 분석가 양성과정의 운영사례를 분석하여 빅데이터 교육 운영에 실천적 지침을 제공하고자 한다. 본 연구는 연구방법으로 사례연구를 채택하였다. 연구의 대상은 교육과정에 참여한 훈련생 26명, 강사 8명, 운영자 1명으로 이들을 대상으로 만족도 설문조사와 인터뷰를 진행하여 데이터 분석 실무능력 함양을 위한 실습 위주의 교과목을 운영하고 장기 실습과정의 원활한 운영을 위한 가이드를 개발할 필요가 있다고 제안하였다[4], [5], [6]. 이와 같은 이유로 대학을 비롯한 많은 기관에서 데이터 사이언스 및 AI 융합인재양성을 위해 학생 융합 시스템을 구축하고 있으나 어떠한 교육이 필요한지, 수강생에게 필요한 교육이 무엇인지에 대한 연구가 절대적으로 부족하다. 따라서 본 논문에서는 K대학교에서 개최된 “2019 빅데이터 청년인재 양성 과정” 수강생에게 교육 참가자의 만족도와 요구를 파악해서 교육과정을 개선하기 위한 기초설문과 과정에 대한 설문을 바탕으로 상관분석을 실시한 후 회귀 분석을 실시하였다.

2. 교육과정 내용 및 방향

K대학교에서는 데이터 사이언스, AI 융합 기반의 수요자 중심, 즉 학생 중심의 교육 및 문제 해결 중심의 MSG의 자기 주도 공동 프로젝트 및 연구, 협업(다양한 분야의 학부생 + 대학원생 협업) 등을 통해 학생들의 역량 강화와 지역 사회에 기여할 수 있는 지속적인 교육, 리서치, 개발, 취업, 창업 등 사회 진출을 위한 솔루션을 제공하는 교육을 실시하였다. 교육대상은 만 34세 이하 미취업 청년을 대상으로 하였으며 선발요건과 교육과정은 다음과 같다.

2.1 선발 요건

데이터 사이언스 및 AI 융합 기반의 빅리더 양성 과정을 이수할 수 있는 기본적인 역량이나 데이터 사이언스 및 AI에 관심을 가지고 본 과정에 적극적으로 임할 학생을 선발하였다.

- R과 Python의 기초 입문 과정 마치었거나 기초 소양을 가지고 있는 학생
- 데이터 사이언스 및 AI 분야로 사회 진출을 희망하는 학생
- 데이터 사이언스 AI 융합 기반의 지역 사회 문제 해결에 관심이 있는 학생
- 선발 방식 : 1차 간단한 본인 소개서 및 지원서 등 서류 접수, 2차 최종 면접을 통해 선발 예정
- 선발 인원 및 비율 : 선발 인원 70명, 교내 학생 50%, 교외 학생 50%

2.2 교육과정

기초 입문 단계(학교 예산으로 진행)에서 본 과정 참여 및 저변 확대를 위한 기초 입문과정은 서울 및 수도권외의 대학과 달리 데이터 사이언스 및 AI 관련 기초 교육을 받을 기회가 적은 지방의 특수성을 감안하였다. 대학에서 본 과정 참여 및 저변 확대를 위한 기초 입문 과정(7월에 실시한 본 집체 교육을 위한 사전 선행 과정)을 진행하였다. 많은 학생이 R과 Python의 기초 프로그래밍 입문을 할 수 있는 과정을 통해 이들 학생 중에서 일부 인원을 선발하였다.

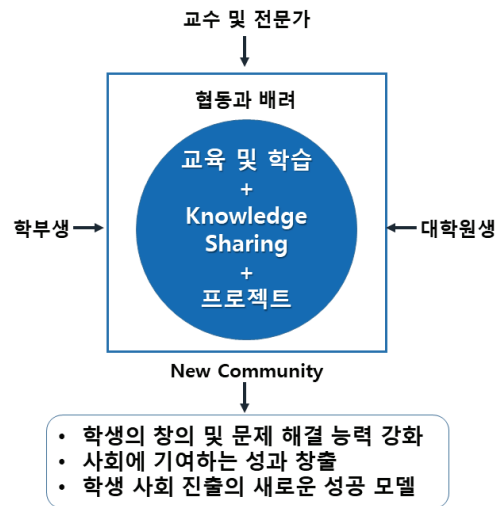


Fig. 1 Composition of big data training

R과 Python 기초 프로그래밍 입문 단계 진행 : 초보자가 배우고자 하는 의지와 열정이 있으면 따라 올 수 있도록 완전 기초 과정으로 진행하였으며, 7월에 실시한 본 교육 과정에 참여하는 학생들 간의 실력 편차를 최소화하여 실질적인 성과를 가지도록 하였다. 본 과정에서는 데이터 사이언스 AI 융합 기반의 빅리더(빅데이터 융합 미래 리더) 양성 과정을 운영하였으며, Real Learning, Real

Impact를 위한 강의와 실습 그리고 배운 내용을 실제로 실습하는 2단계로 프로젝트를 진행하였다. 그리고 Data Science와 AI 융합 과정과 MSG PROJECT과정을 진행하였다.

빅데이터 양성교육의 구성은 Fig. 1과 같이 데이터 사이언스 및 AI 융합 기반의 빅데이터 융합 미래 리더 양성 과정은 학생들을 위한 자기 주도형의 새롭고 창의적이며 효율적인 시스템으로 뜻을 같이 하는 전문가와 교수와 같이 다양한 전공의 학부생과 대학원생을 대상으로 강의와 스터디 그리고 프로젝트를 통한 학생의 실력과 경쟁력 확보 → 창의 및 학습 → 관심분야 / 전공 분야 연구 → 사회 문제를 탐색하고 해결책을 모색 → 프로토타입 제작 및 적용 → 기존 학교 교육에서 머무르지 않는 문제 해결 능력 배양 → 새로운 사회적 가치 창출 → 사회 진출을 위한 성공의 교두보 역할을 하는 방향으로 진행하였으며, R과 Python 기초 교육, Crawler 교육 (Web Crawling 제작과 IoT 센싱 데이터 Crawling), 다양한 예측과 분류를 할 수 있는 기계 학습 그리고 Deep Learning 그리고 AI 챗봇 제작에 이르는 과정을 교육과 실습 과정으로 진행하였다. R과 Python 기초 교육, Crawler 교육 (Web Crawling 제작과 IoT 센싱 데이터 Crawling), 다양한 예측과 분류를 할 수 있는 기계 학습 그리고 Deep Learning 그리고 AI 챗봇 제작에 이르는 과정을 교육과 실습 과정으로 진행하였다.

통상적으로 데이터 사이언스 및 AI 과정을 진행을 하는 경우 학생들 간의 실력 편차가 있어, 중도에 이탈하거나 혹은 포기하는 학생이 발생하는 경우가 상존한다. 이를 해결할 수 있는 실천적 대안이 동료 학습(Peer Learning) 방법이라고 할 수 있다. 그러므로 학생들의 실질적 참여와 만족도를 높이기 위해서 동료 학습을 할 분위기를 조성하고 실력 및 경쟁력 강화를 위한 시너지가 발

생할 수 있도록 하는 것이 필요하였다. K대학에서 진행되는 과정은 교육 및 실습 과정에서 한 단계 나아가 다양한 전공의 학생들이 3중심(자기 주도형 학습 중심, 문제 해결 중심, 협업 중심)으로 진행을 함과 동시에 교수 및 전문가들에게 배우는 것만 아닌 동료들과 협업을 통해 학습과 창의가 이루어 질 수 있는 동료 학습이 이루어질 수 있도록 처음부터 팀을 구성하여 전 과정에 있어서 Peering Learning으로 진행되었다.

미국 Chicago 대학 Data Science for Social GOOD Academy는 실데이터를 사용하여 데이터, 과학자를 양성하는 교육, Social GOOD 프로젝트, Conference 등으로 이루어져 진행되고 있는 미국을 대표하는 명실상부한 Data Science Academy로서 백악관, 월드뱅크, Feeding America 등 행정부, 지자체, 경찰청, 공공 기관 시민 단체 등 각 계각층의 실데이터를 가지고 분석, 예측하는 Social Good 프로젝트를 진행하는 것으로 유명하다.

Fig. 2와 같이 지역 사회적으로 이슈화된 분야의 문제점을 점검하여 학생, 교수, 전문가와 함께 데이터 분석 및 해결 방안을 연구 제시하고자 하였으며, 이러한 연구를 통해 학생의 참여율을 높



Fig. 2 MSG Project planned Area

Table 1. Basic survey

질 문 내 용		인 원	백분율
본 사업 외 인재양성 사업의 참여경험	있 다	13	21.3
	없 다	48	78.7
프로그램의 효율성에 대한 평가	매우 그렇다	23	37.7
	그렇다	32	52.5
	보통	5	8.2
	그렇지 않다	1	1.6
	매우 그렇지 않다	0	0
본 과정에서 가장 관심을 가진 분야(중복응답)	빅데이터	7	11.5
	AI	7	11.5
	크롤링	31	50.8
	딥러닝	15	24.6
	머신러닝	29	47.5
	파이썬	21	34.4
	R	13	21.3
	시각화	3	4.9
	프로젝트	3	4.9
	데이터 사이언스	1	1.6
	IoT	3	4.9
	프로그래밍	2	3.3

Table 2. Basic statistics of data used for analysis

변수구분	질문내용	참여경험	인원	평균	표준편차
종속변수	프로그램의 효율성에 대한 평가	있 다	13	4.38	0.65
		없 다	48	4.23	0.69
		합 계	61	4.26	0.68
독립변수	수강관련 내용에 대한 만족도	있 다	13	4.46	0.66
		없 다	48	4.04	0.77
		합 계	61	4.13	0.76
	수업 또는 직무 연결성에 대한 만족도	있 다	13	4.38	0.65
		없 다	48	4.10	0.78
		합 계	61	4.16	0.76
	자기계발에 도움정도	있 다	13	4.46	0.66
		없 다	48	4.40	0.74
		합 계	61	4.41	0.72

Table 3. Correlation analysis results

변 인	1	2	3	4
1. 프로그램의 효율성에 대한 평가	1			
2. 수강관련 내용에 대한 만족도	.671**	1		
3. 수업 또는 직무 연결성에 대한 만족도	.756**	.684**	1	
4. 자기계발에 도움정도	.631**	.541**	.612**	1

** $p<.01$

Table 4. Regression analysis results

변 인	B	t	R ²	F
(상수)	.741	2.033 *	.646	34.621***
수강관련 내용에 대한 만족도	.212	2.147 *		
수업 또는 직무 연결성에 대한 만족도	.412	3.897 ***		
자기계발에 도움정도	.211	2.175 *		

* $p<.05$, *** $p<.001$

Table 3은 본 인재 양성사업 프로그램의 효율성에 대한 평가, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도간의 상관분석을 실시한 결과이다. 프로그램의 효율성과 가장 큰 상관을 보인 변인은 수업 또는 직무 연결성에 대한 만족도($r=.756$, ** $p<.01$)였고, 다음으로 수강관련 내용에 대한 만족도($r=.671$, ** $p<.01$), 자기계발에 도움정도($r=.631$, *** $p<.01$)로 나타났는데 모두 유의한 정(+)상관관계가 있어, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도가 높을수록 프로그램의 효율성에 대한 평가가 긍정적인 것을 알 수 있었다.

Table 4는 본 인재 양성사업 프로그램의 효율성에 대한 평가를 종속변수로, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도를 독립변수로 설정한 후 회귀분석을 실시한 결과이다. 이 모형은 유의수준 0.1%이하 수준에서 통계적으로 유의한 것으로 나타났다($F=34.621$, $p<.001$), 모형의 설명력은 64.6%

로 나타났다. 또한 수강관련내용에 대해 만족도($p<.05$), 수업 또는 직무 연결성에 대한 만족도($p<.001$), 자기계발에 대한 만족도($p<.05$) 모두 프로그램 참여의 효율성에 통계적으로 유의한 영향을 주는 변인인 것으로 나타났다.

4. 결론

최근 교육계에서 이슈가 되고 있는 빅데이터 양성 교육에서 교과과정 개선을 위한 회귀분석에 따른 만족도 조사에 관한 연구로 결과는 다음과 같다.

첫째, 빅데이터 관련 설문은 지방에 소재한 대학에서 학생 중심의 교육을 실시하였으며, 가장 관심이 있는 분야는 크롤링 50.8%, 머신 러닝 47.3%, 파이썬 34.4%의 순으로 크롤링과 러닝머신에 관심이 매우 높은 것으로 나타났다.

둘째, 교육 프로그램의 효율성에 대한 평가, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도간의 상관분석을 실시한 결과는 프로그램의 효율성과 가장 큰 상관을 보인 변인은 수업 또는 직무 연결성에 대한 만족도($r=.756$, $**p<.01$)였고, 다음으로 수강관련 내용에 대한 만족도($r=.671$, $**p<.01$), 자기계발에 도움정도($r=.631$, $**p<.01$)로 나타났는데 모두 유의한 정(+)상관관계가 있어, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도가 높을수록 프로그램의 효율성에 대한 평가가 긍정적인 것을 알 수 있었다.

셋째, 교육 프로그램의 효율성에 대한 평가를 종속변수로, 수강관련내용에 대해 만족도, 수업 또는 직무 연결성에 대한 만족도, 자기계발에 대한 만족도를 독립변수로 설정한 후 회귀분석을 실시한 결과, 유의수준 0.1%이하 수준에서 통계적으로 유의한 것으로 나타났고, 모형의 설명력은 64.6%로 나타났다. 또한 수강관련 내용에 대한 만족도($p<.05$), 수업 또는 직무 연결성에 대한 만족도($p<.001$), 자기계발에 대한 만족도($p<.05$) 모두 프

로그램 참여의 효율성에 통계적으로 유의한 영향을 주는 변인인 것으로 나타났다.

참고문헌

- [1] 정도준, “빅데이터를 활용한 종합적 재난안전 대책 마련의 필요성,” 국토연구원, 420, pp.108-109, (2016).
- [2] 강영은, 신지영, 박창석, “텍스트마이닝을 활용한 우리나라 기후변화 리스크 평가 및 적응 대책 논의,” 한국도시설계학회지 도시설계, 17(2), pp. 69-84, (2016).
- [3] 이홍주, “한국의 사회적 이슈 도출을 위한 뉴스 빅데이터 분석 연구,” 한국전자거래학회지, 24(3), 163-182(2019).
- [4] 정승화, 도재우, “빅데이터 분석가 양성과정 운영사례 연구,” 교육문화연구, 25(5(B)), pp. 621-640, (2019).
- [5] 김정숙, “빅 데이터 활용과 관련기술 고찰. 한국콘텐츠학회지”, 10(1), 34-40, (2012).
- [6] 정은한, 김경일, “빅데이터시대의 회계교육 과정 개선방안 연구”, 융합정보논문지, 8(5), 145-152, (2018).

(접수: 2019.10.21. 수정: 2019.11.22. 게재확정: 2019.12.05.)