

# 교육 데이터와 분석 기법: 사례 연구를 중심으로

## Education Data and Analytics: A Review of the State of the Art

권영옥

숙명여자대학교 경영학부

### 요약

지난 십여 년간 학생의 학습 과정에서 생성되는 방대한 데이터를 다양한 분석기술을 적용하여 학생의 학습 성과 및 교육 환경을 개선하기 위한 관련 연구들이 활발히 진행되어 왔다. 본 논문에서는 교육데이터의 분석 결과를 성공적으로 활용하고 있는 대학의 사례들을 살펴본 후, 기존 사례 연구를 통해서 보다 구체적으로 분석 목적별로 어떤 데이터와 분석 기법이 사용되는지 정리하였다. 그리고 고찰 결과를 토대로 기존 분석의 한계점 및 향후 분석 방향에 대해 제안하고자 한다.

■ 중심어 : 교육 데이터, 학습 분석, 교육 데이터 마이닝, 학습 성과 예측

### Abstract

With the increase of education data, there have been many studies on the application of various analytics to improve students' performance and educational environments over the past decade. This paper first introduces the cases of universities that successfully utilize the analysis results and, more specifically, examines which data and analytical techniques are used for each analysis purpose. Based on the findings, the limitations of the current analytics and the direction of future analysis are discussed.

■ Keyword : Education Data, Learning Analytics, Educational Data Mining, Student Performance Prediction

## I. 서론

빅데이터, 인공지능, 사물 인터넷 등 새로운 기술의 출현은 기업과 일상생활 뿐 아니라 교육 분야에도 영향을 미치고 있다. 인공지능 기반의 로봇 튜터가 학습에 도움을 주는 것은 이미 현실화 되었으며, 디지털 교과서 또는 학사 시스템을 통해 학생별로 차별화된 맞춤형 교육을 받

을 수 있게 되었다. 미래 학자에 의하면 [1], 기술 진보로 인해 학습이 언제 어디서든 이루어질 수 있으며, 2030년까지 현존하는 직업의 50%가 사라질 것이기에 교육의 변화를 가져올 것이라고 예측한다. 즉, 이러한 변화는 더 이상 선택하고 거부할 수 없는 문제가 아니며, 각 교육 기관에 맞는 방식으로 다양한 첨단 기술을 잘 활용하여 시대에 맞는 교육 서비스를 제공할 수 있

는 방안을 모색해야 한다. 특히, 지난 십년 간 각 대학에서는 축적되어 있는 방대한 디지털 데이터를 학생의 학습 효율성과 교육 환경 개선을 위해 활용하고자 다양한 노력을 기울이고 있다 [2, 3]

본 연구는 교육 데이터의 활용 사례들을 살펴보고, 분석에 사용되는 기법들에 초점을 맞추어 기존 사례 연구들을 고찰하였다. 교육 데이터의 분석은 크게 학습 분석 (Learning Analytics), 교육관리 분석 (Academic Analytics)으로 나뉘는데 각각의 분석에 대해 자세히 알아보고, 실제 대학의 사례들도 살펴본다. 그리고, 보다 구체적으로 사례 연구에서 사용되는 교육 데이터와 분석 기법을 정리하여 소개한다. 예를 들어, 학생의 다양한 학습 활동 데이터를 이용하여 학습 성취도를 예측하는 것이 가장 대표적인 분석의 예라고 할 수 있는데, 성취도에 영향을 주는 요인으로 어떤 데이터를 사용하였으며 어떤 분석 기법을 적용하였는지 살펴본다.

그리고 고찰 결과를 토대로 분석의 한계점 및 향후 분석 방향에 대해 제안하고자 한다. 심리학 분야에서 이미 1950년대부터 “Clinical Versus Statistical Prediction” 를 통해 단순한 예측 알고리즘이 의료 진단, 학생 선발, 야구 선수 영입 등 폭넓은 분야에서 전문가의 판단을 능가한다는 것을 보여주고 있다 [4]. 이는 데이터를 이용한 분석 결과가 전문가의 인사이트보다 확률적으로 도움이 되는 경우가 더 많다는 것을 의미하는데 이에 대해서도 논의한다. 마지막으로, 교육 데이터 분석 및 활용에 있어 국내외 사례의 차이를 살펴보고 국내외 교육 환경의 고려점에 대해서도 알아보하고자 한다.

## II. 데이터 활용을 기반으로 한 교육의 변화

### 2.1 빅데이터와 교육 분석

교육 기관에서 생성되는 다양한 학습, 교육 데이터를 분석하여 활용하는 교육 분석(Education Analytics) 분야는 크게 학습 분석(Learning Analytics)과 교육관리 분석(Academic Analytics)으로 나눌 수 있다 (1st International Conference on Learning Analytics and Knowledge, Banff, Alberta, Feb. 27 - Mar. 1, 2011). 학습 분석은 학생의 학습 효율성 향상을 목적으로 하는 반면, 교육 관리 분석은 비즈니스 인텔리전스 시스템을 이용하여 교육 환경을 효율적으로 관리할 수 있도록 교육 기관 운영 및 지역, 국가 단위 정책 개발을 위한 것이다. 각 분석에 대해서 좀 더 세부적으로 살펴보면 다음과 같다 [5].

#### 2.1.1 학습 분석 (Learning Analytics)

학습 분석이 주목을 받게 된 이유로는 학습과 관련된 데이터의 증가, 온라인 학습 (Open Educational Resources, Massive Open Online Courses 등), 그리고 교육의 성과를 측정하고 보여주고 향상시켜 사회와 그 구성원에게 도움을 주고자 하는 국제적, 국가차원의 관심 증가에 기인한다고 할 수 있다 [6].

학습 분석은 학생의 학습 과정에서 발생하는 다양한 데이터를 실시간으로 분석하고 각 학생에게 맞는 최적화된 학습 모델을 구축함으로써 결과적으로 학습 시스템을 개선할 수 있게 된다. 빅데이터 분석을 위해 사용되는 모든 기술을 적용할 수 있으며, 특히 온라인 학습을 통해 데이터 생성과 수집이 용이해짐에 따라 보다 정교한 분석과 모델링이 가능해졌다.

학습 분석은 학습 활동과 내용을 살펴보고, 어떤 요인, 환경이 학습에 영향을 미치는지 어떻게 개선되어야 하는지 연구한다. 다음과 같이

크게 세 분야로 나누어 설명할 수도 있다. 학생이 학습을 위해 어떤 도움이 필요한지 예측하고, 학습 데이터의 패턴 관계를 분석하며, 학생간 또는 학생과 교수와의 관계를 분석하여 그 영향을 살펴본다. 학습 데이터는 대부분 학습 성과나 학습량과 같은 정형 데이터이나, 학생의 에세이, 교수와의 대화 등과 같은 비정형 텍스트 데이터도 함께 분석할 수 있다.

학습 분석의 단계를 나누어 보면 다음과 같다 [7]. 가장 기본적인 분석이라 할 수 있는 특정 과목의 학습 과정을 분석하는데 나아가 소셜 네트워크 분석을 통해 학생의 사회적 관계를 분석하기도 한다.

다음 단계는 교육 데이터 마이닝 (Educational Data Mining)을 이용한 예측 분석이다. 이전에는 교육 데이터 마이닝을 학습 분석과 구분하였으나 최근 들어 학습 분석의 일부로 포함하여 설명하고 있다. 이 단계에서는 데이터 마이닝 기법을 활용하여 통계적인 패턴이나 규칙, 알고리즘, 모델을 찾는 것을 넘어 다양한 예측 모델을 적용하는데 초점을 두고 있다.

본 연구에서는 주로 이 단계에서 사용되는 분석 기법을 사례를 통해 살펴보고자 한다. 학생의 인구 통계학적 특성, 과거 학습 성과, 온라인 로그인 횟수, 토론 참여도 등과 같은 데이터를 분석하여 향후 수강하게 되는 과목의 학업 성과나 활동에 대해 예측한다. 예측 결과는 자기 주도형 학습을 효율적으로 진행할 수 있도록 도와주며, 학습 과정에 교수가 효율적으로 개입하여 가이드를 줄 수 있도록 돕는다.

다음 단계로는 학생의 학습 활동이나 성과를 바탕으로 해당 학생에게 최적화된 학습 콘텐츠, 즉 지능형 또는 적응형 교육 콘텐츠 (intelligent or adaptive content)를 제공하는 것이다. 마지막 단계는 학생의 수준과 학습 스타일에 맞게 학습 정보나 방법, 피드백 등을 제공하는 적응형 학습 (adaptive learning)이다.

## 2.1.2 교육 관리 분석(Academic Analytics)

교육 관리 분석은 학습자체의 개선보다는 기관 시스템의 개선에 더 초점을 둔다. 다양한 교육 운영과 관련된 데이터를 이용하여 조직이나 제도 운영을 효율적으로 만드는 것을 목적으로 한다. 조직 단위에서의 분석은 조직성과의 평가, 지식 관리를 들 수 있고, 지역이나 국가 단위에서는 조직의 운영 정보를 통합하여 교육 정책을 수립하는데 활용될 수 있도록 한다.

앞서 언급하였듯이 기업에서 데이터 기반의 의사결정을 위해 사용하는 비즈니스 인텔리전스의 개념을 교육 기관에 적용한 학문, 활동을 의미한다. 교육 관리 분석을 통해 관련 행정 업무를 자동화하고 행정 업무에 대한 부담을 줄일 수 있다면 교육의 질을 높이는데 더 많은 시간을 투자할 수 있게 될 것이다. 예를 들어, 입학 관련 업무, 반복적인 서류 처리 업무 등을 줄이고 생산적인 학습행동 증진을 도와줄 수 있게 된다.

## 2.2. 맞춤형 교육을 통한 혁신 대학 사례

### 2.2.1 Arizona State University

지난 십년간 많은 대학들이 다양한 기술을 접목하여 교육 데이터를 활용하고 있다. 그 중 가장 대표적인 사례가 최근 몇 년간 혁신적인 대학 1위로 선정된 애리조나 주립대학이다 [8]. 어느 대학보다 교육 데이터를 활용하여 학생들의 학업을 성공적으로 마칠 수 있도록 돕는데, 데이터 분석 결과에만 따르는 데이터 기반 (data-driven)이 아닌 데이터와 함께 다른 조건들을 고려하여 의사결정을 하게 되는 데이터를 정보로 활용하는 (data-informed) 방식임을 강조하고 있다.

애리조나 주립대의 대표적인 데이터 활용 시스템은 온라인 상에서 상담을 해주는 'eAdvisor'이다. 학생들이 자신의 관심 있는 분야와 전공

을 찾을 수 있도록 도움을 주고 수강 신청 시에도 가장 효율적인 수업과 시간을 조언해준다. 전공별 수강 과목의 순서도 제안하여 특정 전공을 위해서 가능한 저학년때 수강해야 하는 기초 과목들도 알려준다. 학생들은 가이드라인으로 주어진 트랙에서 벗어났을 때 바로 도움을 받을 수 있다. eAdvisor를 통해 졸업률은 11.6% 향상되었으며 학생유지율(Retention)은 84%까지 향상되었다. 또한, 학습을 도와주기 위해 다양한 시스템은 도입하였는데 McGraw Hill Education이 개발한 ALEKS 시스템과 GogBooks가 개발한 적응학습 시스템이 그 사례이다. ALEKS를 이용하여 학생들의 수학 학습 능력을 평가하고, 차별화된 학습 방법을 제공함으로써 Algebra 과목을 성공적으로 이수할 수 있도록 하였다.

데이터 활용 이외에도 전체적인 교육 과정도 혁신적으로 산업 협력, 지역 밀착형 교육으로 변화를 주었다. 모든 강의를 프로젝트 중심으로 변경하였다. 학생들이 지역 기업 또는 지역 사회의 문제점을 해결하는 프로젝트를 수행하도록 하였다. 또한, 69개의 학과를 폐지하고 필요에 따라 학과를 합쳐 융합학부를 만들기도 하였다.

1학년 학생들은 보다 광범위한 분야 접할 수 있게 하고, 관심 있는 세부 전공을 탐색할 시간을 가질 수 있도록 하였고, MOOC 플랫폼도 적극적으로 활용하고 있다. MIT에서 개발한 edX의 대학교 1학년 과정을 이수하면 애리조나 주립대에 2학년으로 입학할 수 있다(Global Freshmen Academy), 이러한 프로그램은 다른 대학에서 보기 어려운 혁신적인 것이다.

### 2.2.2 Georgia State University

조지아 주립대학교는 학생들의 성공을 방해하는 요인들을 찾을 수 있는 데이터 기반의 반복적인 방법을 이용하였다 [9]. 대규모의 데이터 웨어하우스를 구축하고 기관의 혁신을 적절할

때에 조금씩 범위를 늘려가면서 효과적으로 이를 수 있도록 하였다.

학업 성취도가 낮은 과목에 대해서는 이전에 좋은 성적을 받은 학생들로부터 개인 교습을 받을 수 있도록 하는 등 추가적인 도움을 준다. 기초 수학 과목에 대해서는 면대면 또는 컴퓨터 기반의 지도를 받을 수 있도록 커리큘럼을 재설계하였다.

경제적인 문제 때문에 중도 탈락하게 될 위험이 있는 학생들에게 장학금을 지급하도록 하였다. Graduation Progress System (GPS)을 개발하여 예측 분석 기법을 적용하여 중도 탈락의 요인을 분석하였다. 학생이 돌아올 수 있도록 하기 위해서 지도교수가 언제 학생들을 만나야 하는지 알려주는 조기 알림 시스템도 개발하였다. 학부 상담을 하나의 단위로 통합하고 상담자의 수를 두 배 이상으로 늘렸다. 1학년 학생들에게 입학 전 여름 학기에 수업을 듣고 학업 및 재정 상담을 받을 수 있도록 하여 첫 학기를 성공적으로 보낼 수 있도록 도와준다.

또한, 유사 전공을 모아 놓는 메타 전공(meta-major)을 만들어 1학년 학생들이 세부 전공을 정하기 전에 비슷한 관심사를 가진 다른 학생과 함께 수강할 수 있도록 하였다.

조지아 주립대는 행정적인 면에서도 다양한 시도를 하였다. 재정 지원, 학업 지원, 상담, 입학, 수강신청 등의 업무를 모두 결합시키도록 행정 구조를 변경하였다. 6년 만에 졸업하는 비율이 2003년 32%에서 2014년 54%로 증가하였다.

## III. 교육 데이터와 분석 기법

### 3.1 분석 목적과 기법

2008년에서 2013년 사이에 진행된 209개의 학습 분석과 교육 데이터 마이닝에 관한 연구를

선정하여 분석한 결과는 다음과 같다 [2]. 연구 분야, 학습 목적, 데이터 수집, 분석 기법과 결과 등에 따라 연구를 구분하여 분석하였고, 교육 데이터 분석의 목적은 크게 다음의 6가지로 나눌 수 있다.

- 학생의 행동 모델링: 학생의 학습 패턴과 성과의 관계를 찾아서 각 학생의 학습 전략을 모델링 한다.
- 학업 성과 예측: 성과에 영향을 미칠 것이라 생각되는 변수를 몇 개 더 추가한다고 해서 예측의 정확도가 높아지지 않는다. 답을 맞힐 확률이 가장 높거나 가장 낮은 극단적인 경우를 제외하고 분석한 경우 예측력을 높일 수 있다.
- 자기반성과 인식: 분석 결과를 보고 어떤 학생이 어떤 도움이 필요한지 인지하고 찾아낼 수 있어야 한다. 학생 스스로도 다른 학생과 비교하여 자신의 학습 진도 및 상황을 판단할 수 있어야 한다.
- 학생 유지/이탈 예측: 학생 만족도, 인지된 교육의 유용성, 효율성 등을 이용하여 분석 가능하다.
- 피드백과 평가 서비스: 분석 결과는 의미 있는 피드백을 주어 학생 스스로 인지하고 반성할 수 있는 기회를 가질 수 있도록 해야 하며, 이는 학생의 이탈 의도를 낮추는데 도움이 된다.
- 리소스 추천: 학습에 필요한 리소스를 추천하는데, 일반적인 추천 알고리즘을 이용한다.

사용된 데이터 분석 방법으로는 분류, 클러스터링, 회귀, 텍스트 마이닝, 연관 법칙, 시각화, 소셜네트워크 분석, 통계 등을 고르게 사용한 것을 알 수 있다 [2]. 2002년부터 2015년까지 이루어진 연구 중 학생의 성과 예측에 사용된 기법을 살펴보면, Decision tree, Artificial Neural

Networks, Naive Bayes, K-Nearest Neighbor and Support Vector Machine 으로 크게 나눌 수 있다 [10].

### 3.2 교육 데이터와 추천 알고리즘

e-Commerce에서 사용자가 가장 좋아할 만한 제품이나 서비스를 추천해 주는 것과 같은 원리로 추천 알고리즘을 교육 데이터에 적용 가능하다.

학생의 과목별 성과를 예측할 때에도 기존의 다른 분야에서 널리 사용되는 추천 기법을 적용할 수 있다. 예측 결과를 이용하여 학생마다 차별화된 맞춤 지원을 하고 이는 결과적으로 학생의 성과를 높일 수 있다. 2011년부터 2017년까지의 연구를 살펴본 결과 [11], 대표적인 추천 알고리즘인 Collaborative Filtering, Matrix Factorization, Restricted Boltzmann Machines, Personalized Multi-Linear Regression Models 등을 이용하여 학생의 학업 성과를 예측하였다.

성과 예측 뿐 아니라 학업 상담, 학습, 강의를 추천할 수 있다 [12]. 제품이나 서비스 대신에 교과목, 학습 자료, 학습 주제, 학업 성과 등의 데이터를 이용한다. 또한, 전공과 부전공을 추천해 줄 수 있다. Park [13]은 Collaborative Filtering 과 Matrix Factorization 기법을 이용하여 학점을 예측하고 가장 적합하다고 여겨지는 전공과 부전공을 추천하는 프로세스를 제안하였다. 여러 전공/부전공의 필수, 선택 과목 중 수강하지 않은 과목의 성적을 예측하고 평균적으로 가장 높게 예측된 전공 리스트와 특히 좋은 성적을 얻게 될 과목들을 보여준다. 그리고 학생들은 전공/부전공 선택 또는 변경시 추천 결과를 참조하게 된다. 입학 후 전공 선택을 해야 하거나, 입학시 선택한 전공에 대해서 확신이 서지 않거나, 혹은 부전공 선택시 도움이 필요할 경우 학생 뿐 아니라 지도교수와 학교 행



정 관리자에게 가이드를 줄 수 있는 유용한 툴로 사용될 수 있다. 성적 외에 관심사, 성격, 졸업 후 진로 계획을 포함하면 더 좋은 결과를 얻을 수 있을 것이다.

### 3.3 성과 기반 분석 사례

조지 메이슨 대학과 미네소타 대학의 Computer Science and Engineering 전공 학생들의 성적데이터와 LMS 데이터, 그리고 Stanford University MOOC 데이터를 이용하여 학업 성과를 분석하였다 [14].

#### 3.3.1 학습 성취도

이전에 수강한 과목들의 학점을 학생-과목의 매트릭스로 표현하는데 학생  $i$ 의 과목  $j$ 에 대한 학점은  $g(i,j)$ 로 나타낼 수 있다. 교과목 수에 비해 한 학생이 수강하는 과목의 비율은 매우 낮으므로 앞에서 생성한 매트릭스는 밀도가 낮다. 다음 두 가지 기법을 이용하여 학점을 예측하였다.

- Regression-based methods

회귀 기반의 기법을 이용하여 한 학생이 특정 과목에서 얻게 될 학점을 예측하였다. 이러한 분석은 기존에 수강한 과목과 다음 학기에 수강할 과목이 내용면에서 연계되어 있다고 가정한다. 즉, 선수 과목으로 반드시 들어야 하는 과목들을 대상으로 분석하였다.

학생별로 개인화된 회귀 모델을 구축하였는데, global bias, student bias, course bias를 통제하며 학업 성적에 영향을 주는 요소로는 학생 관련 (skill, motivations, learning habits), 과목 관련(교과목 자료, 난이도 등)을 포함하였다.

- Matrix factorization-based methods

학생과 과목을 k-dimensional latent feature vector로 나타내고 특정 학생의 특정 과목에 대

한 성적은 두 벡터의 곱으로 계산된다. 앞서 언급한 바와 같이 교과목들이 선수과목이 있는 것을 전제로 하기 때문에, 관련 교과목들에 특화된 모델을 수립하였다. 추가적인 학생 정보(나이, 인종, 성별, 지역, GPA, SAT 점수, 고등학교 성적, 이전 학기 까지 취득한 학점, 수강 과목 내용, 과목 담당 교수)를 고려하기 위해서는 일반적인 Matrix Factorization 모델을 이용해야 한다.

#### 3.3.2 LMS와 MOOC에서의 학습활동

Learning Management Systems (LMSs)은 교수와 학생간의 의사 소통의 툴로 사용되는 온라인 플랫폼으로 여기에서 축적되는 데이터를 이용하여 학생의 학업 단계를 추적하고 미래의 학업 성적을 예측할 수 있다. Stanford University MOOC 데이터는 통계 한 과목에 대한 클릭스트림 데이터만을 수집하여 이용하였다. 클릭스트림 데이터로는 세션, 퀴즈, 비디오, 과제, 시간 데이터를 포함한다.

LMS 상에서의 학생의 활동 여부를 예측하는데, 다음 3가지 요소를 이용한다.

- 학업 성과: 현재 GPA, 해당 과목의 이전 활동(과제 또는 퀴즈)의 결과
- LMS 활동과 과목 특성: 활동 타입 (과제와 퀴즈), 과목 난이도, 과목이 개설된 학과
- LMS 특성: 시스템 상에서 학생들의 활동을 나타내는 서버 로그 데이터, 포럼에 글을 올린 횟수, 포럼의 내용 및 수업 자료 열람 여부, 이전 활동에 대한 성적 등을 포함한다.

LMS 또는 MOOC에서의 학습 활동을 통하여 퀴즈와 같은 수업 시간 내 평가와 과제에 대한 성과를 예측한다. 예측결과는 궁극적으로 학생들이 적성에 맞는 전공을 선택할 수 있도록 도울 뿐 아니라, 도움이 필요한 경우 상담자 또는

교수에게 주의 메시지를 보내어 초기에 조치를 취할 수 있도록 한다.

### 3.3.3. 학생 유지 (Retention)

학업 성취도 뿐 아니라 대학에서 주요한 척도 중 하나는 학생 유지율이다. 특히 온라인 대학과 MOOC의 강의는 저렴한 비용으로 수강의 기회는 많으나 중도에 포기하지 않고 끝까지 수강하는 학생의 확률이 상대적으로 매우 낮다. 따라서, 온라인 학습이 이루어지는 경우 특히 학생의 유지율을 높이는 것이 매우 중요하다.

앞에서 사용된 성과 기반의 예측 방법을 이용하여 (performance-prediction techniques) 어떤 학생이 중도탈락 하거나 수업에서 낙제할 것인지 예측할 수 있다. 또한 이전에 수강한 과목에서의 활동 데이터와 분석하고자 하는 과목의 몇 주 전 데이터를 통합하여 분석함으로써 실시간 예측도 가능하다 [15].

### 3.3.4 취업 및 진로 조언

국내 교육의 특성상 학업 성과 예측과 더불어 취업 성과 예측에 대한 연구가 활발히 진행되고 있다. 대학평가의 기준이 학생의 취업관련 역량에 목표를 두고 있기 때문에, 온라인 학습 활동보다는 취업률에 영향을 미치는 요인을 파악하고자 한다.

기본적으로 대학의 지명도와 소재 지역이 취업성과에 중요한 요인으로 분석되었다 [16]. 같은 대학의 최근 3년간의 졸업생을 대상으로 한 연구에서는 [17], 취업성과를 사회경제적 요인(전공계열, 성별, 신입학/편입, 정원내/외, 모집 유형, 출신고교 지역, 출신고교 유형), 학교생활 요인(졸업평점, 교내성적장학금 수혜, 교환학생, 복수전공, 경상계열 복수전공, 의무/초과학기 등록), 그리고 취업준비 요인(인턴십, 취업지원 프로그램 참가)으로 나누어 분석하였다. 로지스틱회귀분석과 의사결정나무를 사용하였고,

취업에 미치는 영향이 인턴십 여부, 취업지원 프로그램 참가여부, 전공계열, 졸업평점평균, 경상계열 복수전공 여부, 성별, 입학구분 등의 순으로 높게 나타났다.

사회경제적 요인 중 영향력이 큰 전공계열, 성별, 입학구분의 경우 취업을 준비하는 학생들이 변경을 할 수 없는 요인이지만, 졸업평점평균, 경상계열 복수전공 여부, 인턴십 여부, 취업지원 프로그램 참가여부 등은 개인의 노력으로 변경할 수 있는 요인이므로 이러한 분석 결과를 취업 지도에 이용할 수 있을 것이다. 또한, 취업에 성공한 졸업생들의 학사 정보를 이용하여 유사한 분야에 취업하고자 하는 학생들에게 진로 조언을 할 수 있으며, 대학 내부의 데이터 뿐 아니라 외부 데이터까지도 활용할 수도 있다 [18].

## IV. 한계점 및 향후 연구 방향

앞서 살펴보았듯이 방대한 교육 데이터를 분석하고 활용할 수 있는 방안에 대한 연구는 활발히 진행되고 있다. 하지만, 분석 결과를 적용하고 실행하는 것은 또다른 문제이다. 교수 또는 상담자의 개입(intervention)유형에 따라서도 다른 결과를 가져올 수도 있을 것이다. 따라서, 실행 방법에 따른 학생의 학습 태도, 행동, 인지 에 미치는 영향에 대한 연구 또한 필요하다 [19].

앞의 대학 사례에서 볼 수 있듯이 교육 데이터를 잘 활용하고 있는 대학은 이미 학업 성과, 유지율, 졸업률에서 큰 성과를 보여주고 있다. 이에 비해 국내 대학에서는 학습 분석에 대한 관심이 저조한 편이다. LMS의 사용도 수업 자료를 전달하는 것 이외에 다른 기능들을 잘 활용하고 있지 못하다. 그리고 대학 내에서 MOOC 플랫폼도 보다 적극적으로 활용할 수 있는 방안도 모색해 보아야 할 것이다.

또한, 학생의 학업 성과를 예측하는 다양한 분석 기법을 이용하기 위해서는 예측 알고리즘

에 대한 신뢰가 전제되어야 할 것이다. 데이터가 충분하지 못한 경우도 있을 수 있으며 예측력에 대한 판단 기준 또한 상황에 따라 다르기 때문에, 예측 알고리즘은 학생 또는 교육자의 의사결정을 자동화하는 것이 아닌 도와주는 데 그 목적이 있다고 할 수 있다. 알고리즘을 이용하는 것이 효과적인 업무와 교육자가 잘 하는 것을 구분하여 조화를 이루어야 한다. 이와 더불어, 데이터 분석 결과는 특정 기관 내 학생의 학습과 교육 환경을 개선을 위한 것으로 알고리즘 혹은 모델 개발도 교육 기관의 특성이 반영되도록 설계되어야 할 것이다.

본 연구에서는 데이터 분석 기법으로 데이터 마이닝을 중심으로 살펴보았는데, 향후에는 더 많은 진보된 기술들을 활용할 수 있을 것이고 이에 대한 학습 성과에 미치는 영향에 대한 분석이 필요할 것이다. 게임 기반 학습, 모바일 학습, 뇌과학 기반의 학습 활동 등이 전통적인 학습 방법에 비해 효과적인지 비교 분석해 볼 수 있을 것이다. 마지막으로, 기존 연구들이 주로 학습자 중심의 연구였다면 교육자를 위한 데이터 활용방법에 대해서도 보다 심도 있는 연구가 필요할 것으로 보인다.

## 참 고 문 헌

- [1] Frey, T. *Communicating with the Future: How Re-engineering Intentions Will Alter the Master Code of Our Future*, Da Vinci Institute Press, 2011.
- [2] Papamitsiou, Z. and Economides, A.A. "Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence," *Journal of Educational Technology & Society*, Vol.17, No. 4, pp.49-64. 2014.
- [3] 권영욱. "빅데이터를 활용한 맞춤형 교육 서비스 활성화 방안연구. 한국지능정보시스템학회 학술대회논문집, pp. 90-96, 2013.
- [4] Meehl, P.E. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. MN, US: University of Minnesota Press. 1954.
- [5] 정윤혁, 빅데이터와 교육분석 (Education Analytics). 미디어와 교육, 제5권, 제1호, pp.44-49, 2015.
- [6] Ferguson, R. "Learning analytics: drivers, developments and challenges," *International Journal of Technology Enhanced Learning*, Vol.4, No.5/6, pp. 304-317, 2012.
- [7] Siemens, G. and Long, P. "Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, Vol.46, No.5, pp. 31-40, 2011.
- [8] <https://blog.simonassociates.net/what-makes-arizona-state-university-the-most-innovative-college-in-the-us>
- [9] <http://www.sr.ithaka.org/publications/building-a-pathway-to-student-success-at-georgia-state-university>.
- [10] Shahiri, A. M. and Husain, W. "A review on predicting student's performance using data mining techniques," *Procedia Computer Science*, Vol.72, pp. 414-422, 2015.
- [11] Iqbal, Z., Qadir, J., Mian, A.N. and Kamiran, F. "Machine learning based student grade prediction: A case study," arXiv preprint arXiv:1708.08744, 2017.
- [12] Manouselis, N., Drachler, H., Verbert, K., and Santos, O.C. (Eds.). *Recommender Systems for Technology Enhanced Learning: Research Trends and Applications*. Springer Science & Business Media, 2014.
- [13] Park, Y. "A Recommender System for Personalized Exploration of Majors, Minors, and



Concentrations,” *In RecSys Posters*, 2017.

[14] Elbadrawy A, Polyzou A, Ren Z, Sweeney M, Karypis G., and Rangwala H. “Predicting student performance using personalized analytics,” *Computer*, Vol.49, No.4, pp:61-69, 2016.

[15] Boyer, S. and Veeramachaneni, K. “Transfer Learning for Predictive Models in Massive Open Online Courses,” *In International conference on artificial intelligence in education*, pp. 54-63, 2015.

[16] 최기성, 조민수 “대학 명성이 졸업생 취업 질에 미치는 효과와 시사점”, *조사연구*, 제17권, 제2호, 119-162, 2016.

[17] 염동기, 문상규, 박성수. “대학졸업자의 취업성과 결정요인에 관한 실증분석,” *취업진로연구*. 제7권, 제4호, pp.45-68, 2017.

[18] 김종율, 노광현. “AI 를 활용한 대학생 진로 조언 시스템 모델 및 데이터 수집과 융합에 대한 연구,” *디지털융복합연구*, 제17권, 제2호, pp. 177-185, 2019.

[19] Rienties, B., Cross, S., and Zdrahal, Z. Implementing a learning analytics intervention and evaluation framework: What works?. *In Big data and learning analytics in Higher Education* (pp. 147-166). Springer, Cham, 2017.

## 저자 소개



### 권영옥(YoungOk Kwon)

- 연세대학교 컴퓨터과학과 (이학사)
- 서울대학교 경영학과 (석사)
- University of Minnesota, Carlson School of Management, Information and Decision

Sciences (박사)

- 현재 : 숙명여자대학교 경영학부 부교수
- 관심분야 : Personalization, Analytics for education/business/healthcare, analysis of IT industry and firms