

역량중심 교육과정 개발을 위한 학업성취도 예측 시스템: D대학 사례를 중심으로

Learning Ability Prediction System for Developing Competence Based Curriculum: Focusing on the Case of D-University

김성국¹, 오창현^{2*}

¹두원공과대학교 IT융합학부, ²한국기술교육대학교 전기전자통신공학부

Sungkook Kim¹, Chang-Heon Oh^{2*}

¹Div. of IT Convergence, Doowon Technical University, Anseong 17520, Korea

²School of Electrical, Electronics and Communication Engineering, KOREATECH, Cheonan 31253, Korea

[요약]

대학에서의 학업성취도란 대학교육을 통한 결과로서 학생들이 구현한 질적 변화와 발달의 수준이라는 포괄적 의미로 인식되고 있다. 따라서 대학생의 학업성취도는 창의성, 리더십, 글로벌 역량 등 다양한 인재상에 대한 시대적, 사회적 요구와 연계되어 그 의미를 부여하게 되지만 실질적으로 대학교육의 성과지표로서 중요하게 인식되고 있는 것은 학점으로 귀결되고 있다. 이러한 학점을 통한 학업성취도의 측정은 많은 문제를 가지고 있는데, 특히, 평가 방식과 내용 그리고 대학의 서열화 효과 등에 의해 학점을 통한 학업성취도의 표준화는 매우 어려운 문제로 인식되고 있다. 본 연구는 머신러닝 기법을 활용하여 D대학 졸업생을 대상으로 학업성취도의 우수 여부를 예측하는 시스템을 제시한다. 사용된 변수는 일부 개인정보와 졸업연도, 학번, 학과명, 계열명 등의 학사 정보 등 최대 96개를 활용하여 분석하였으나 개인정보나 학과정보 등은 이미 결정되어 노력에 의해 변경될 수 없는 데이터므로 분석 대상이 될 항목은 이미 결정된 데이터를 제외한 학과별/학생별 역량으로 한정하였다. 본 연구에서는 경기권 소재 전문대학인 D대학의 미션, 비전, 교육목표 및 인재상 등이 반영된 핵심역량의 분석을 통해 학업 성취도 예측 시스템을 구현해 보고, 해당 시스템의 도입이 학업성취도에 미치는 영향을 머신러닝을 활용하여 예측하기 위해 진행되었다. 향후 연구결과를 학과에서 진행되는 교육과정 수립 및 학생 지도 등에 적용하여 학업성취도를 향상시킬 수 있는 근거를 마련하는데 활용할 예정이다.

[Abstract]

Achievement at university is recognized in a comprehensive sense as the level of qualitative change and development that students have embodied as a result of their experience in university education. Therefore, the academic achievement of university students will be given meaning in cooperation with the historical and social demands for diverse human resources such as creativity, leadership, and global ability, but it is practically an indicator of the outcome of university education. Measurement of academic

<http://dx.doi.org/10.14702/JPEE.2022.267>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 20 July 2022; Revised 17 August 2022

Accepted 24 August 2022

*Corresponding Author

E-mail: choh@kut.ac.kr

achievement by such credits involves many problems, but in particular, standardization of academic achievement by credits based on evaluation methods, contents, and university rankings is a very difficult problem. In this study, we present a model that uses machine learning techniques to predict whether or not academic achievement is excellent for D-University graduates. The variables used were analyzed using up to 96 personal information and bachelor's information such as graduation year, department number, department name, etc., but when establishing a future education course, only the data after enrollment works effectively. Therefore, the items to be analyzed are limited to the recommended ability to improve the academic achievement of the department/student. In this research, we implemented an academic achievement prediction model through analysis of core abilities that reflect the philosophy, goals, human resources image, and utilized machine learning to affect the impact of the introduction of the prediction model on academic achievement. We plan to apply the results of future research to the establishment of curriculum and student guidance conducted in the department to establish a basis for improving academic achievement.

Key Words: Academic ability, Big data, Core competency, Curriculum, Machine Learning, Prediction system

1. 서론

대학교육의 성과로서 학업성취도는 일반적으로 대학의 비전, 미션과 교육목표가 교육과정을 통하여 대학생들에게 전달되는 과정을 의미한다. 즉, 대학에서의 학업성취도란 대학교육에 의하여 만들어진 결과로서 학생 개개인의 학업적 수준의 향상이라는 의미로 인식 되고 있다[1]. 따라서 대학생의 학업성취도는 리더십, 글로벌, 창의력, 윤리 등 대학의 다양한 인재상에 대한 시대적, 사회적 요구와 연계되어 도출하고 있다. 그러나 아직까지도 실질적으로 대학교육의 성과 지표로서 중요한 역할을 차지하고 있는 것은 학점으로 구성된 학업성취도임을 부인할 수 없다. 이와 같이 학점에 의한 대학생들의 학업성취도 측정은 여러 문제점을 가지고 있으며 평가 방식과 대학의 서열화 등에 의해 학점에 의한 학업성취도를 표준화하여 비교하는 것은 어렵기 때문에 외국의 경우 학점 대신 ACT, GRE와 같은 표준화된 측정결과를 사용하고 있다[2,3]. 그럼에도 불구하고 우리나라의 학업성취도에 대한 연구는 아직도 일반적인 척도로서 학점을 대부분 사용하는 것을 볼 수 있다.

이러한 대학의 교육 결과를 판단하는 지표로 사용되어 온 학점에 대한 신뢰도는 점점 낮아지고 있다. 대학알리미의 2022년도 공시자료에 의하면 전국 4년제 220개 대학교 졸업생 평균학점이 3.45이고 A학점이 40.5%, B학점이 52.6%로 나타나 졸업생의 93%가 평균 B학점 이상을 취득한 것으로 조사되어 있다. 전문대 135개 대학의 경우, 졸업생의 평균학점이 3.49이고 A학점 이상이 35.2%, B학점 이상이 52%로 나타나 졸업생의 87.25%가 평균 B학점 이상을 취득하고 있는 것으로 나타나 있다[4]. 이와 같이 대학의 학점은 상대평가로 진행됨에도 불구하고 학점이 높은 학생들이 많은 소위 학점 인플레이션이 존재하고 있는 것을 잘 알고 있다.

얼마전까지 대학정원이 고등학교 졸업자 수를 만족하지

못하였기 때문에 대학입시는 치열했으나 몇 년 전부터 상황이 조금씩 변화하고 있다. 대학의 문이 유례없이 넓어지고 있어 대학의 정원이 고등학교 졸업생 수 보다 많아지고 있어 이제는 누구나 대학에 입학할 수 있는 상황이 되었다. 이는 자연스럽게 대학생들의 기초학력 저하로 이어지게 되었다. 이러한 대학생들의 기초학력 저하 문제는 이전에 비해 기초학력이 부족하고 수업 준비도 부족한 학생들이 대학에 진학하게 되면서 발생하게 된 일반적인 현상이 되었다. 그러나 대학의 지원은 대부분 일회성 프로그램 운영이 대부분이며 학내 부처별로 독립적으로 운영되는 한계를 가지고 있다. 관련 연구도 이공계열 학생을 대상으로 한 기초수학에 대한 연구가 주로 이루어지고 있어 다양한 전공의 학생들을 배려하고 지원하는 데는 한계가 있다[5].

선행 연구를 보면 이러한 전문대학생들의 기초학습 저하 원인을 정서적, 인지적, 행동적 측면으로 보고 있다. 먼저 정서적 측면에서 보면 충동성, 사회 부적응, 낮은 성취동기 및 학습에 대한 흥미 결여가 있고 인지적 측면에서 보면 낮은 기억력과 학습전략의 부족, 집중력 부족, 사고력 결여 등을 들고 있다. 마지막으로 행동적 측면에서는 쓰기 및 문제해결 역량의 부족과 지각 장애 등을 들고 있다[6,7].

따라서 요즘 대학교육의 핵심 중 하나는 학업 성취도가 낮은 학생들에 대한 교육의 질 제고에 있다. 본 연구는 핵심역량 중심 교육을 진행하고 있는 D대학의 사례를 통해 핵심역량과 학업성취도와의 상관관계를 분석하여 학과별, 학기별로 중요한 핵심역량을 추천하여 교육과정에 적용시켜 학생들의 학업성취도를 향상시키는 것을 그 목적으로 한다.

본 연구는 D대학 졸업생의 학업성취도를 향상시키고자 기존에 운영중인 핵심역량 교육과정을 머신러닝 기법을 활용하여 우수 학업성취도 예측모델을 개발하고 그에 맞춰 핵심역량 중심 교육과정을 개발·운영하는데 있으며 연구의 내용은 기존 사례 분석, 분석방법 모델링 및 결과 도출을 통한

모델 설계로 구성되어 있다. D대학의 핵심역량 중심 교육과정은 학생들에게 필요한 핵심역량을 도출하고 이를 향상 시키기 위해 여러 교육 프로그램을 활용하여 교육과정을 개발하고 운영하고 있다. 본 연구에서는 D대학의 핵심역량 중심 교육과정 개편이 재학생의 학업성취도 향상에 어느정도 기여했는지를 머신러닝 기법을 활용하여 예측하고자 한다.

II. 기존 연구 분석

A. 핵심역량의 개념

최근 대학교육에 있어서 핵심역량의 중요성이 높아지고 있으며, 각 대학별로 핵심역량에 기반한 교육과정 개편이 진행되고 있다. 특히 대부분의 대학들은 기존의 NCS 교육과정에서 핵심역량 중심 교육과정으로의 전환을 진행하고 있으며, 교육부에서도 역량 중심 교육으로의 변화를 지원하고 있다. 이런 변화의 바탕에는 현재의 체제에서는 미래 인재를 양성하기 어렵다는 위기감이 자리하고 있으며, 전문 분야에 대한 지식과 기술의 습득과 여러 지식과 상황을 해결할 수 있는 문제 해결 능력으로써 역량에 주목하게 된 것이다[8].

일반적으로 핵심역량이란 대학생활 및 조직생활을 성공적으로 수행하는데 필수적인 능력으로[9]업무를 수행하는 곳에서 우수한 성과를 낼 수 있는 개인의 능력이라는 의미에서 “조직의 성과를 효율적으로 높이는데 필요한 구성원의 지식, 스킬, 태도 등의 조합”으로 정의한다[10]. 특히 최근에는 대학교육에 많이 활용되고 있어 전공분야에 대한 지식, 리더

십 및 자기주도성 등 대학생들에게 요구되는 역량과 글로벌 능력, 문제해결 능력, 윤리의식, 직무 및 직업 소양 등 예비 직업인으로서 요구되는 역량이 강조되고 있다[11].

교육부에서는 이러한 상황을 반영하여 2006년부터 한국 직업능력개발원과 대학교육의 성과관리 차원에서 대학들의 교육역량을 지속적으로 모니터링하기 위하여 “대학생 핵심역량 진단 시스템(K-CESA: Korea Collegiate Essential Skills Assessment)”을 개발하였다. 현재 국내에서 사용되는 핵심역량에 대한 진단검사는 다양하지만, 대학생의 핵심역량을 진단하기 위한 진단 도구는 부족한 실정이며, K-CESA는 각 대학의 다양한 인재상에 맞는 핵심역량을 측정하기에는 많은 문제점이 발생하고 있다.

D대학은 이러한 K-CESA의 문제점을 해소하고자 자체적으로 개발을 진행하였으며 해당 역량 모델은 “James D. Klein”이 제시한 역량 모델 개발 방법론[12]을 응용하여 2019년 2월, 대학 내부의 의견 수렴과 미션, 비전 및 인재상 등과의 연계를 통해 그림 1과 같이 수립되었다[13].

이를 살펴보면, 대학 내·외부 환경분석 및 의견수렴을 통해 “교육혁신으로 창의·융합·인성 및 실무능력을 겸비한 전문 직업인 양성”을 대학 미션으로 설정했고 이를 구체화시키고자 “지역사회 기반의 창의·인성을 중시하는 미래지향적 이공계 중심대학”이란 비전을 제시하였다. 비전을 달성하기 위해 책임감 있는 인재, 실무형 인재, 창조적 인재의 3개 인재상을 수립 하였고 이를 다시 리더십, 글로벌, 성실, 의사소통, 기초학습 등 10개의 핵심역량으로 세분화 하여 제시하였다. 이처럼 D대학의 핵심역량은 대학의 미션과 비전, 그리고 인재상에 맞춰 그에 필요한 역량을 도출하였다. 이렇게 수립된



그림 1. D대학 핵심역량 체계도

Fig. 1. D-University Core Competency System Diagram.

표 1. D대학교의 핵심역량 정의[13]

Table 1. D-University Core Competency Definition

역량명	정의
기초학습	읽기·쓰기·셈하기 등의 학습 기초적 능력
창의력	예술적, 과학적, 디자인적 사고, 기업가 정신 등 새로운 생각을 해내는 힘
의사소통	생각이나 감정 등을 교환하는 개방적이며 유연하고 명료한 사고 및 협상 능력
비판적 사고	자료 조사·분석·이해·평가 및 정보 활용능력 등의 사고 과정
협업	다수 노동자가 상호협력하는 작업형태이지만, 21세기에서는 지식·아이디어를 공유·수용·적용하는 공간을 만드는 능력
문제해결	기존 방식을 깨는 고찰·조사·선택·평가·체계화·정보 등의 해석 능력
글로벌	국제사회에 공통적으로 통용되는 가치·준칙·제도·관행 등 보편적 가치를 수용하는 능력
성실	열린 사고·용기·호기심·열정·자기 주도·투지의 참된 성품형
리더십	책임감·조직력·조정력·협상력·술선수법·대인관계의 현신적 영웅형
윤리	자기인식·자기실현·자존감·박애·정직·의리·사랑·정의감·충성심의 仁義禮智德 모형

10개의 핵심역량을 표 1과 같이 도출하여 2019년 3월부터 핵심역량 교육을 운영하고 있다.

B. 머신러닝의 개념

머신러닝(Machine Learning)은 기계학습 이라고도 하며, 인공지능의 한 분야로 수집된 대량의 데이터를 이용하여 정의되지 않은 패턴을 학습하여 결과를 만들어내는 학문으로 데이터, 패턴인식, 컴퓨터 연산 등이 합쳐진 분야이다. 다시 말해 머신러닝은 컴퓨터가 수집된 데이터로부터 패턴을 학습하여 결과를 예측하는 모델을 만들고, 생성된 모델을 평가,

개선시키는 것으로 정의할 수 있다. 현재 머신러닝은 음성 및 영상인식, 이미지 처리 등의 분야에 적용되어 기존 데이터의 패턴이나 미래의 일을 예측하는 핵심기술로 활용되고 있다.

머신러닝을 설명할 수 있는 몇 가지 개념들을 바탕으로 해당 모델의 정확성 및 유용성을 판단할 수 있다. TP(True Positive), FP(False Positive), FN(False Negative), TN(True Negative)의 개념으로 TPR, FPR, Accuracy, Recall, Precision, F-measure, ROC area의 의미를 해석하여 예측모델의 성능을 분석하게 된다. 머신러닝 성능평가를 위한 개념들의 상세한 의미와 수식은 표 2와 같다[14].

표 2. 머신러닝 성능평가 용어[14]

Table 2. Machine learning performance evaluation terms

용어	의미	표현식 또는 판단기준
TP rate	관측값이 참인 경우 중 예측값도 참인 비율을 의미	$\frac{TP}{(TP+FN)}$
FP rate	관측값이 거짓인 모든 값들 중 예측값이 참인 비율	$\frac{FP}{(FP+TN)}$
CA(Accuracy)	전체 예측값 중 관측값과 정확하게 일치하는 비율	$\frac{TP+TN}{(TP+TN+FP+FN)}$
Recall	관측값이 참인 경우 중 예측값이 참인 비율(정확하진 않더라도 관련된 대부분이 분류되는지 여부)	$\frac{TP}{(TP+FN)}$
Precision	예측값이 참인 모든 값들 중 관측값이 참인 비율(얼마나 정확하게 분류하는지 여부)	$\frac{TP}{(TP+FP)}$
F1(F-measure)	Precision과 Recall의 조화평균을 의미	$\frac{TP}{TP+0.5(FP+FN)}$
AUC	예측결과와 실제 결과 간의 예측성능 비교정보, 곡선 아래의 면적이 클 수록 예측 성능이 좋은 것으로 판별	- AUC<=0.5 : 유효성 가치 없음 - 0.5<AUC<=0.7 : 유효성이 낮음 - 0.7<AUC<=0.9 : 유효성 중간 - 0.9<AUC<1 : 유효성이 높음 - AUC=1 : 유효성 완벽

머신러닝은 컴퓨터가 인공지능을 얻기 위해 학습하는 알고리즘으로 총 세 가지 유형이 존재한다. 먼저 지도학습은 정답값이 있는 데이터셋으로 학습하는 방식으로 컴퓨터가 답을 구한 후 정답과 비교해 나가며 오차를 줄여가는 학습방식이다. 이는 답이 주어지기 때문에 오차를 구하기 쉽고 성능 측정도 용이한 장점이 있다. 지도학습은 회귀와 분류가 주로 다루어 지는데 회귀는 실수 값을 예측하는 문제로 답을 구하는 함수를 구하는 방식이다. 분류는 데이터가 속하는 그룹을 예측하는 문제로 그룹을 나누는 함수를 구하는 방식이라고 정의된다. 이러한 지도학습 알고리즘은 대표적 예시로 선형 회귀, 로지스틱 회귀, CART, KNN 등이 있다.

두번째로 비지도학습은 정답값이 없는 데이터셋으로 학습하는 방식으로 주로 입력 데이터만 있고 출력 변수가 없는 경우에 사용되고 있다. 비지도학습은 정답이 없기 때문에 데이터의 패턴을 찾아내서 그 결과가 의미가 있다면 사용하고 의미가 없다면 파라미터를 조정하여 다시 학습하는 과정을 반복적으로 진행하는 방식이다. 클러스터링 방식이 비지도 학습 알고리즘에 해당된다. 지도학습과 비지도학습의 가장 큰 차이점은 데이터 레이블링에 있다. 지도학습의 경우 학습 데이터가 어떤 데이터인지 미리 구분해줘야 하지만 비지도 학습의 경우 구분되지 않은 데이터들 안에서 일정한 구조를 도출해 가는 과정을 진행한다.

마지막으로 강화학습은 행동심리학에서 영감을 받아 알고리즘화 되었으며, 현재 상태에서 어떤 행동을 하는 것이 적합한지 학습하는 방식이다. 일련의 어떤 행동을 할 때마다 보상이 주어지는데, 강화학습은 보상을 최대화 해주는 방향으로 학습이 이루어진다. 강화 학습은 훈련 데이터가 제시되지 않으며, 오류에 대해서도 수정을 하지 않는다는 점에서 일반적인 지도 학습과 다르다.

이러한 머신러닝 기술도 분명히 한계가 있다. 우선 알고리즘상에서 데이터를 통해 예측의 기초 패턴을 파악하므로 입력 데이터가 명확하지 않을 경우 컴퓨터는 최적화된 추측만 가능하고 논리적인 단계를 상식적으로 추론하는 부분에 있어서는 추측이 어려운 문제가 있다. 또한 사람의 언어를 이해하거나 번역과 음성을 인식하는 등의 기능은 현재 기술상 머신러닝 능력의 한계가 극명하게 보이며 일반인의 개인정보의 침해 등 관련 법률과 사회 규범이 제대로 갖춰지지 않은 상태에서 오는 유동성도 불안요소 중 하나이다. 마지막으로 머신러닝의 알고리즘상의 보편화를 위한 성차별, 인종차별 등 부정적이고 고정화된 관념을 포함한 인류의 부적절한 측면을 학습할 수 있다는 점에 특히 주의해야 한다.

III. 분석대상 및 변수

A. 분석대상

본 연구의 데이터는 경기도 소재 D대학 졸업생들의 개인 정보, 입학정보 및 학사정보 등을 취합한 자료이다. 2019년 2월부터 2021년 2월까지 졸업한 학생이 대상이 되었다. 2022년 2월 졸업생에 대한 데이터는 아직 취합 되지 않아 생략되었다. 향후 이를 보완하여 추가로 연구를 진행할 예정이다. 본 연구에 사용된 데이터는 표 3과 같이 종속변수인 학업성취도 우수 여부(Y/N)가 포함된 37개 학과의 5,278명 졸업생 관련 정보이다. 2019년 2월 졸업자는 36.4%(1,922명)이고, 2020년 2월 졸업자는 33.6%(1,771명), 2021년 2월 졸업자는 30%(1,585명)이다. 이는 다시 안성 캠퍼스가 61.4%(3,241명)이고 파주 캠퍼스가 38.6%(2,037명)로 조사되었다. 계열별로 보면 기계자동차계열이 24.4%(1,288명), IT계열이 16.6%(874명), 간호보건계열이 10.7%(566명), 인문사회계열이 16.2%(856명)로 조사되었다.

표 3. 변수의 빈도표

Table 3. Variable frequency table

구분	변수명	빈도수	비율
졸업연도	2019	1,922	36.4%
	2020	1,771	33.6%
	2021	1,585	30.0%
캠퍼스	안성 캠퍼스	3,241	61.4%
	파주 캠퍼스	2,037	38.6%
성별	남성	3505	66.4%
	여성	1,773	33.6%
계열	IT계열	874	16.6%
	간호보건계열	566	10.7%
	군부사관계열	101	1.9%
	기계자동차계열	1,288	24.4%
	디스플레이공학계열	450	8.5%
	디자인계열	340	6.4%
	방송계열	248	4.7%
	인문사회계열	856	16.2%
	자연과학계열	350	6.6%
	호텔관광계열	205	3.9%
학업성취도 우수여부	Y(평균 90점 이상)	1,217	23.06%
	N(평균 90점 미만)	4,061	76.94%

B. 변수 설명

본 연구는 D대학의 핵심역량 중심 교육과정을 수료한 졸업자들이 우수한 학업성취도를 갖추도록 교육과정을 수립하여 교육의 질 향상을 기하는데 있다. 따라서 주소지, 출신고 등의 개인정보나 학과명, 계열명과 같이 본인의 노력에 상관없이 이미 결정된 변수는 제외하고 분석을 진행하였다.

목표변수는 학업성취도의 우수 여부로 Y/N의 이항변수로 설정하였다. 우수 여부는 졸업평균점수가 90점 이상일 경우 “Y”로 미만일 경우 “N”로 구분하였다. 이는 일반적으로 대학에서 우수한 성적을 “A”학점 이상으로 보고 있으며 학칙을 통해서도 교과목 당 “A”학점을 획득하는 학생의 수를 제한하는 등 엄격히 관리하고 있기 때문이다. 본 연구의 기준이 되는 D대학의 경우 A학점 이상을 해당 교과목 수강자의 40%미만으로 학칙에 규정되어 있다. 위의 표 3을 보면 목표변수의 빈도는 분석대상 5,278명 중 학업성취도 우수자는 1,217명(23%)이고, 비우수자는 4,061(76.94%)이다.

일반적인 통계분석 방식과 동일하게 머신러닝 분석에서도 변수가 반드시 필요하다. 본 연구에서도 D대학 졸업자의 학업성취도에 영향을 줄 수 있는 다양한 변수가 포함되어 있다. 기존의 일반적인 통계분석 방법은 여러 변수들의 상호작용을 분석하여 예측 결과를 얻지만 머신러닝 분석에서는 다양한 변수를 포함하더라도 자유도에 문제가 발생하지 않기 때문에 가능한 모든 변수를 분석하여 학업성취도의 우수 여부를 예측해 보고자 하였다. 아래의 표 4는 분석에 사용된 변수를 범주별로 보여주고 있다.

우선 기본적인 변수로서 학사정보 중 졸업연도와 학번, 학과명, 계열명 등을 설명 변수에 포함시켰다. 대학생활 동안의

성과를 종합적으로 평가하는 DW인증제도 관련 변수는 학기별 학점과 일부 중복이 되어 제외하였다. 이 외에 가장 중요한 변수라 할 수 있는 글로벌, 기초학습, 협업, 비판적 사고, 창의력, 의사소통, 문제해결 등 7개 스킬 역량과 성실, 리더십, 윤리 등 3개 인성역량 데이터가 포함되었다. 스킬역량과 인성역량은 각 학기별로 수집되었는데, 학과에 따라 적게는 44개에서 많게는 88개의 데이터가 산출되었다. D대학은 2/3/4학년제가 혼재된 학제를 운영하고 있어 학과별, 학제별로 그에 맞도록 데이터를 수집하여 분석하였다. 즉, 2년제 학과는 학기별/역량별로 44개의 데이터가 수집되었고, 3년제 학과의 경우 66개, 4년제 학과는 88개가 수집 되었다. 또한 교육과정 편성이 학과별로 다르기 때문에 결측치가 예상보다 많았고 특히 마지막학기의 데이터는 현장실습 등으로 NULL값이 많았다. 예측 모델링의 원리는 이전 데이터의 패턴을 학습하여 미래에도 적용될 것이라고 판단하는 것이므로 과거의 데이터가 없다면 그 부분에 대해서는 예측이 불가능하거나 정확도가 낮아진다. 따라서 결측치를 줄인 데이터를 수집하기 위한 노력이 지속적으로 필요할 것으로 보인다.

IV. 결과분석 및 검토

A. Logistic regression model

Logistic regression model은 범주형 종속변수 1개와 여러 개의 독립변수들과의 인과관계를 알아보는 분석방법이다. Logistic regression model은 회귀 계수의 해석이 가능한 점과 과적합 가능성이 적고 특정 분류 모델의 성능을 비교하고 싶

표 4. 사용된 데이터의 범주 및 항목

Table 4. Data categories and items

범주	개수	항목
학사정보	7	졸업연도, 학번, 학과명, 계열명, 캠퍼스, 학점평균1(4.5만점), 학점평균(100점만점)
학기별/역량별 학점	88	1학년 1학기
		1학년 2학기
		2학년 1학기
		2학년 2학기
		3학년 1학기
		3학년 2학기
		4학년 1학기
		4학년 2학기
학업역량 우수여부	1	목표변수
합계	96	

기초학습, 창의력, 비판적 사고, 의사소통, 협업, 문제해결, 글로벌, 성실, 리더십, 윤리, 평균 점수 등 각각 11개 항목
단, 2/3/4년제 학과별로 분석 항목 수 차이 존재

표 5. 생성된 Logistic regression model 평가 결과

Table 5. Logistic regression model evaluation result

model	AUC	CA	F1	Precision	Recall
Lasso(채택)	0.968	0.919	0.948	0.941	0.954
Ridge	0.967	0.917	0.947	0.940	0.953

을 경우 베이스 모형을 활용하여 비교해 볼 수 있다는 장점이 있다. 본 연구에서는 scikit-learn의 “LogisticRegression” 함수를 이용하여 로지스틱 회귀 분석을 모델링 하였다. 규제를 완화하기 위해 C값을 ‘20’으로 세팅하였고 충분한 훈련을 위해 Max_iter값은 ‘1,000’으로 세팅하였다. 또한 소규모 데이터에 적합한 ‘liblinear’ 알고리즘을 사용하였다. 표 5는 해당 모델의 평가 결과이다. Logistic regression model은 정규화 된 정도에 따라 Lasso와 Ridge의 두가지 방식으로 측정하는데 Lasso방식은 회귀계수 절대값의 합이 최소화 되도록 조정하는 방법이고 Ridge방식은 회귀계수 제곱합이 최소화 되도록 회귀계수를 조절하는 방법이다. 본 연구에서는 실험을 통해 정확도와 AUC가 더 높은 Lasso 방식을 채택하여 분석을 진행하였다. 평가 항목인 AUC, CA, F1, Precision, Recall 모두 1에 가까울수록 좋은 모델로 평가된다.

B. Decision Tree model

Decision Tree model은 입력값에 대한 예측값을 트리 구조로 나타내어주는 모델로 입력값에 따른 설명변수의 변화를 직접 볼 수 있기 때문에 해석이 쉽다는 큰 장점이 있는 분석 방식이다. 그러나 상대적으로 다른 지도 학습모형에 비해 예측력이 떨어진다는 단점이 있다. 본 연구에서는 scikit-learn의 “DecisionTreeClassifier” 함수를 이용하여 모델링 하였다. 표 6은 해당 모델의 평가 결과이다. 본 모델은 성장(Growing)과 가지치기(Pruning) 단계를 통해 구현되는데 성장 단계에서는 최적화할 목적함수를 정하고 각 마디에서 목적함수를 최적화하는 변수와 분리 기준을 찾아 분석을 해나가는 알고리즘을 가지고 있다. 가지치기 단계에서는 과적합을 방지하기 위해 해석이 안 되는 규칙 등 불필요한 부분을 제거하는데 이러한 과적합을 막기 위해 부모 마디는 ‘100’, 자식 마디는 ‘50’, max_depth는 ‘100’으로 설정하여 측정하였다.

표 6. 생성된 Decision Tree model 평가 결과

Table 6. Decision Tree model evaluation result

model	AUC	CA	F1	Precision	Recall
Tree	0.944	0.917	0.916	0.915	0.917

표 7. 생성된 Random Forest model 평가 결과

Table 7. Random Forest model evaluation result

model	AUC	CA	F1	Precision	Recall
Random Forest	0.970	0.932	0.931	0.931	0.932

Criterion은 ‘gini’, splitter은 ‘best’로 세팅하였고 그 밖의 매개 변수는 기본값을 그대로 사용하였다. 평가 항목인 AUC, CA, F1, Precision, Recall 모두 1에 가까울수록 좋은 모델로 평가 된다.

C. Random Forest model

Random Forest model은 위에서 설명한 Decision Tree model을 여러 개 모아서 데이터 분류 및 예측을 수행하는 알고리즘으로 어떤 집단에 대한 분류나 예측을 진행할 때, 하나의 의사결정나무를 사용하는 것보다 여러 개를 결합해서 사용하면 보다 높은 성능의 모델을 만들 수 있다는 큰 장점이 있다. 본 연구에서는 scikit-learn의 “RandomForestClassifier” 함수를 이용하여 랜덤 포레스트 분석을 모델링 하였다. 실험결과 매개변수 n_estimators의 값은 정확도에 큰 영향을 미치지 않아 기본값(‘10’)을 그대로 사용하였다. 표 7은 해당 모델의 평가 결과이다. 평가 항목인 AUC, CA, F1, Precision, Recall 모두 1에 가까울수록 좋은 모델로 평가된다.

D. AdaBoost model

AdaBoost model은 근래 캐글에서 우승을 차지하는 가장 인기 있는 모델링 방법 중 하나이다. AdaBoost는 Adaptive Boost의 약자로 초기 모형을 Weak Learner로 지정하고 모든 프로세스마다 가중치를 정하여 이전 모델의 약점을 보완하여 나가는 알고리즘이다. 일반적인 부스팅 모델은 Weak Learner를 이용하여 Strong Learner를 만들어내는 특성을 가지고 있다. 이 모델의 단점은 예측을 제대로 하지 못한 비정상적인 데이터가 존재하는데 이러한 비정상적 데이터를 학습하여 예측력이 강한 모형으로 모델링 된다고 하여 “Adaptive”란 단어가 붙게 되었다. Adaboost 알고리즘은 분

표 8. 생성된 AdaBoost model 평가 결과

Table 8. AdaBoost model evaluation result

model	AUC	CA	F1	Precision	Recall
AdaBoost	0.854	0.896	0.896	0.896	0.896

류 문제, 회귀 문제에 모두 적용 가능하며 이전 모델이 제대로 예측하지 못한 데이터에 더 집중하여 이를 최적화하는 모델을 만들게 되는 특징을 가지고 있다. 본 연구에서는 scikit-learn의 “AdaBoostClassifier” 함수를 사용하여 Adaboost 분석을 모델링 하였다. 실험결과 매개변수 learning_rate의 값은 정확도에 큰 영향을 미치지 않아 기본값(‘1.0’)을 그대로 사용하였고 분류 모델의 개수 n_estimators값은 ‘100’으로 세팅하였으며 알고리즘(solver)은 확률에 기반한 ‘SAMME.R’을 사용하였다. 표 8은 해당 모델의 평가 결과이다. 평가 항목인 AUC, CA, F1, Precision, Recall 모두 1에 가까울수록 좋은 모델로 평가된다.

E. Neural Network model

Neural Network model은 기계학습과 인지과학 분야에서 고안된 학습 알고리즘의 하나로 신경세포인 뉴런의 신호 전달체계를 모방하여 노드로 표현되는 인공의 뉴런이 학습을 통해 결합 세기를 변화시켜 문제를 예측하고 분류하는 알고리즘을 말한다. 인공신경망은 뉴런들이 신호를 받아 임계값을 넘어가면 그것을 전달하는 신경세포의 신호 전달 과정을 모방하여 만들어졌다. 이를 수학적인 모델로 변형한 것이 퍼셉트론인데 이는 각각의 가중치의 크기를 조절하여 입력 신호의 크기를 정하고 이렇게 나온 입력 신호의 합에 활성화 함수를 결합하여 결과값을 얻는 방식이다. 인공신경망에서 예측 결과값을 찾는 과정이 “입력층 → 은닉층 → 출력층”의 순서로 진행되는데 이를 순전파라 하고 가중치의 수정은 반대로 “출력층 → 은닉층 → 입력층”으로 진행되기 때문에 역전파라고 한다. 본 연구에서는 scikit-learn의 “MLPClassifier”

표 9. 생성된 Neural Network model 평가 결과

Table 9. Neural Network model evaluation result

model	AUC	CA	F1	Precision	Recall
Neural Network	0.964	0.920	0.920	0.919	0.920

함수를 이용하여 모델링 하였다. 실험결과 알고리즘(solver)은 ‘sgd’를 사용하였고 규제 of 양을 제어하는 alpha 값은 ‘0.01’로 세팅하였고 그 밖의 매개변수는 정확도에 큰 영향을 미치지 않아 기본값을 그대로 사용하였다. 표 9는 해당 모델을 평가한 결과이다. 은닉층은 1~2개를 사용하고 각각 최대 100개의 뉴런을 사용하였다. 은닉층의 활성화함수는 입력값이 0보다 작으면 0을 출력하고, 0보다 크면 입력값 그대로를 출력하는 정류 선형 유닛 함수를 사용하였고 가중치 최적화 방식은 확률적 경사 하강법을 채택하였다. 평가 항목인 AUC, CA, F1, Precision, Recall 모두 1에 가까울수록 좋은 모델로 평가된다.

표 10. 모델별 학업성취도 우수 여부 예측 결과

Table 10. Prediction result of excellent academic achievement by model

모델	정확도 (%)	ROC/AUC (%)
Logistic regression model	94.1	96.8
Random Forest model	93.1	97.0
Neural Network model	91.9	96.4
Decision Tree model	91.5	94.4
AdaBoost model	89.6	85.4

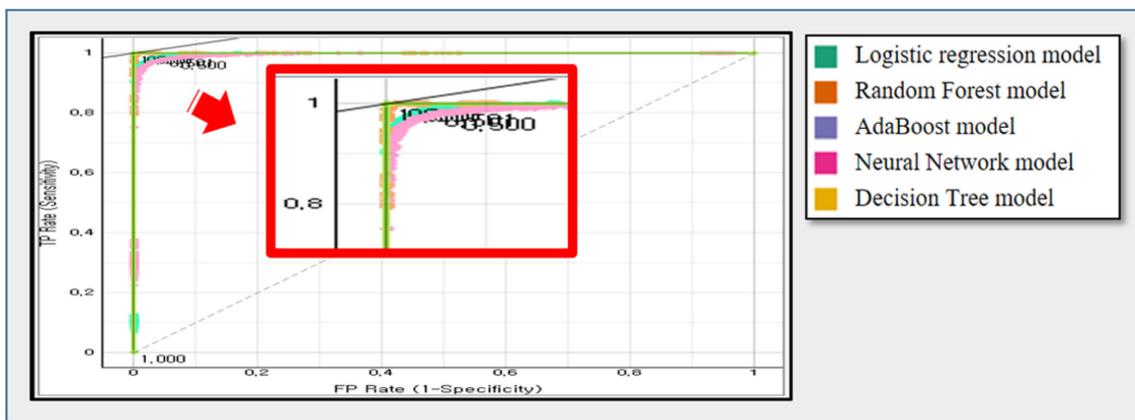


그림 2. 학업성취도 우수 여부 예측 ROC curve

Fig. 2. Excellent academic achievement prediction ROC curve.

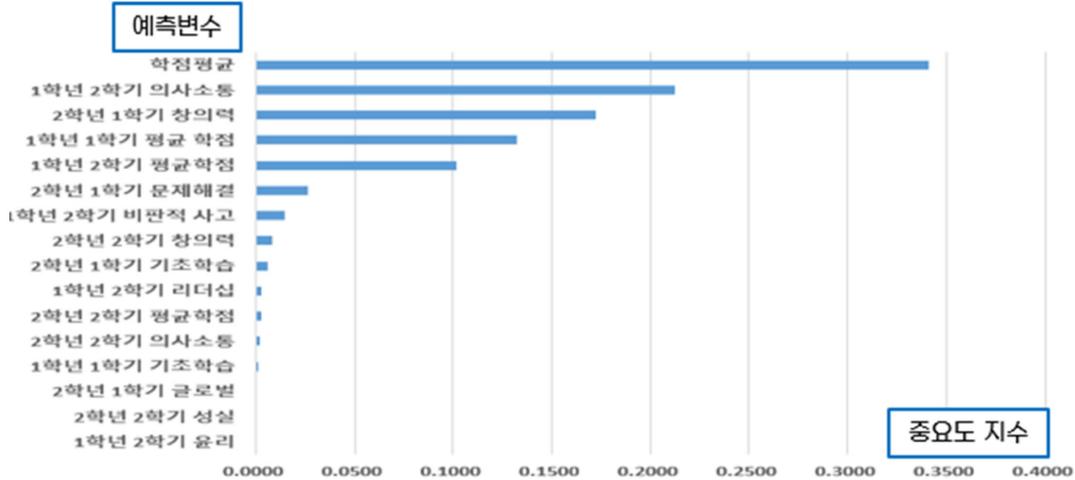


그림 3. 우수 학업성취도 달성 여부 예측 변수 중요도 상위 16개

Fig. 3. Top 16 most important variables for predicting excellent academic achievement.

F. 결과 검토

본 연구의 D대학교 졸업생들의 핵심역량에 대한 취득 학점을 분석하여 학업성취도의 달성여부를 예측하는 모델을 설계하는 것을 목표로 하였고 이를 위해, 종속변수를 핵심역량 중심 교육을 수료한 학생의 학업성취도 우수 여부로 설정하고 각각의 교과목과 매핑된 핵심역량의 평균 점수를 변수로 놓고 다양한 분석 방식을 적용하여 가장 예측력이 좋은 분석 방법을 도출하였다. 그 결과 표 10 및 그림 2과 같이 Logistic regression model이 정확도(94.1%)와 AUC(96.8%)에서 가장 우수한 것으로 나타났고 Random Forest model(정확도 93.1%, AUC 97.0%)이 그 뒤를 이었으며 AdaBoost model(정확도 89.6%, AUC 85.4%)이 가장 미흡한 예측 결과를 보여줬다. 따라서 이후 모델링은 Logistic regression model을 이용하여 실험을 진행하였다.

본 연구의 결과는 그림 3과 같이 조사되었는데 높은 점수

로 나타난 변수일수록 분석모델이 결정에 더 많이 고려했다고 볼 수 있다. 우수 학업성취도 예측에서는 학점평균, 1학년 2학기 의사소통, 2학년 1학기 창의력, 1학년 1학기 평균학점, 1학년 2학기 평균학점의 순서로 중요도가 높았으며 이 순위는 입력된 데이터에 따라 달라질 수 있다. 흥미로운 점은 학점평균이나 1학년 1학기 평균 학점의 중요도가 높게 나타났다는 점이다. 이는 학업성취도의 우수 여부는 결론적으로 학점과 연관성이 많기 때문에 다른 핵심역량에 비해 우수 학업성취도의 예측에 유리하게 작용한 것으로 보인다.

표 11은 “학생별 우수 학업성취도 달성을 위한 역량 추천(예시)”로 각 학과에서 이 자료를 활용하여 학생들의 성적, 취업 및 진로 상담 등에 활용할 수 있을 것이다. 당연한 이야기지만 학생의 노력에 의해 우수한 학업성취도의 달성이 가능하다는 것을 수치화된 데이터로 보여주고 학습을 독려한다면 대학 전체의 학업 성취도 향상에 긍정적인 효과를 기대할 수 있을 것이다.

표 11. 학생별 우수 학업성취도 달성을 위한 역량 추천(예시)

Table 11. Ability recommendation to achieve excellent academic achievement by student

성명	우수 성취도 달성 확률	추천역량1	추천역량2	추천역량3	추천역량4	추천역량5
김XX	99.82%	3학년 2학기 문제해결	4학년 2학기 리더십	4학년 1학기 의사소통	1학년 2학기 윤리	3학년 1학기 리더십
박XX	0.39%	3학년 2학기 문제해결	4학년 2학기 리더십	4학년 1학기 의사소통	3학년 1학기 리더십	2학년 2학기 의사소통
이XX	0.12%	2학년 2학기 비판적사고	2학년 1학기 성실	3학년 2학기 문제해결	4학년 2학기 리더십	4학년 1학기 의사소통
조XX	97.02%	2학년 2학기 의사소통	3학년 2학기 문제해결	4학년 2학기 리더십	4학년 1학기 의사소통	3학년 1학기 리더십

표 12. 학과별 우수 학업성취도 달성을 위한 역량 추천(예시)

Table 12. Ability recommendation to achieve excellent academic achievement by department

학과	우수 성취도 달성 확률	추천역량1	추천역량2	추천역량3	추천역량4	추천역량5
A	81.28%	1학년 2학기 비판적사고	2학년 2학기 의사소통	3학년 1학기 윤리	3학년 1학기 의사소통	3학년 1학기 비판적사고
B	74.41%	3학년 1학기 의사소통	3학년 1학기 비판적사고	3학년 1학기 협업	3학년 1학기 문제해결	2학년 2학기 문제해결
C	79.22%	2학년 1학기 협업	2학년 1학기 문제해결	2학년 2학기 문제해결	2학년 2학기 성실	2학년 1학기 성실
D	71.49%	3학년 1학기 기초학습	3학년 1학기 창의력	3학년 1학기 비판적사고	3학년 1학기 협업	3학년 1학기 문제해결

표 12는 “학과별 우수 학업성취도 달성을 위한 역량 추천 예시”로 학과의 교육과정 개발 시 이를 참고한다면 대학 전체의 학업성취도 향상에 많은 도움이 될 것으로 기대한다.

V. 결론

본 연구는 D대학의 핵심역량 중심 교육과정(CBC: Competency-Based Curriculum)을 이수한 졸업생들의 핵심역량과 학업성취도와의 관련성을 분석하여 D대학의 교육과정 개선 및 대학 교육 정책에 반영되는 것을 주 목적으로 한다.

앞의 실험을 통해 정확도가 94.1%로 측정된 Logistic regression model을 채택하여 적용하였다. 본 실험을 통해 설계된 모델을 테스트 데이터에 적용하면서 모델의 성능이 안정적인 분포를 나타내는지 확인하였다. 데이터는 일반적으로 적용되는 수준인 70%를 학습 데이터로 하고 30%를 테스트 데이터로 사용하였다. 실험 결과를 보면 데이터에 따라 정확도와 ROC/AUC 지표가 변화하기도 하였으나 유의미한 수치가 아님을 확인할 수 있었다.

본 연구를 통해 설계된 시스템은 학업성취도의 우수 여부 예측을 진행하는데, 학과별 학기별로 학업성취도 향상을 위해 집중해야 할 역량을 추천한다. 이를 차년도 교육과정에 적용한다면 해당 학과의 학업성취도 향상에 긍정적인 역할을 할 것으로 보인다. 학생별 우수 학업성취도 달성을 위한 역량 추천을 통해 지도교수는 학생 개개인의 성적관리를 효과적으로 진행할 수 있으며 학생들도 본인들의 우수한 성취도를 획득할 확률을 보고 그에 맞는 노력을 더 할 수 있다는 점에서 본 연구의 의의를 찾을 수 있다. 또한 학과별 우수 학업성취도 달성을 위한 역량 추천 자료를 통해 학과 교육과정을 수립할 때 응용한다면 학과 전체적인 학업성취도의 향상을 가져올 수 있을 것으로 기대하고 있다.

그러나 졸업생에 대해 별도의 구분없이 학업성취도가 우수한지 여부로만 판단하여 여러가지 한계점은 분명히 있었다. 저성취의 요인에 대한 규명이 불가능하다는 것이 가장

큰 문제로서 향후 연구에서는 학업 저성취의 원인을 알아보고 어떻게 극복할 수 있는지도 추가적으로 조사할 예정이다. 또한 코비드19로 촉발된 비대면 수업과 학업성취도간의 연관관계 등도 조사하여 학사운영에 반영한다면 학업성취도의 질도 파악할 수 있을 뿐만 아니라 해당 대학의 핵심역량 중심 교육의 우수성과 적절성도 함께 분석이 가능할 수 있을 듯 하다. 이러한 결과를 활용하여 교육과정을 지속적으로 보강해 나간다면 서두에서 이야기했던 대학생 기초학력 부족 문제의 해결에 많은 도움이 될 것으로 판단된다.

또한 머신러닝의 가장 큰 단점인 설명력을 높이기 위해 본 연구를 통해 추천된 역량을 교육과정상에 보강한 교육이 학업성취도에 영향을 주었는가 하는 효과성 분석과 설명 가능한 인공지능(XAI: eXplainable Artificial Intelligence)을 통한 추가 연구를 진행할 계획이다. 현재까지 머신러닝은 예측 정확도는 향상되었지만 도출 과정상의 타당성 확보나 도출된 결과의 근거에 대한 논리적 설명이 어렵기때문에 신뢰성이 필요로 하는 시스템에서는 머신러닝의 사용이 매우 제한적이었다. 향후에는 XAI 기술을 적용하여 해당 모델이 왜 그런 분석 결과를 내렸는지, 어떤 역량을 발전시키는 것이 학과 혹은 학생의 학습성취도 향상에 필요한지 등의 연구를 진행한다면 예측을 뛰어넘는 보다 더 효과적인 분석도 충분히 가능할 것으로 판단된다.

참고문헌

- [1] A. W. Astin, *What Matters in College?: Four Critical Years Revisited*, Jossey-Bass Publishers, October 2001.
- [2] J. I. Lee and J. H. Kim, “A study on the relationship between college students' essential skills and academic achievement,” *The Journal of Vocational Education Research*, vol. 31, no. 2, pp. 227-246, 2012.
- [3] H. K. Kim and H. S. Shin, “Factors that influence the academic achievement of college students: a preliminary

- study for changing higher education policy,” *The Journal of Educational Administration*, vol. 20, no. 1, pp. 171-188, 2002.
- [4] Higher Education in Korea, Graduation grade distribution of graduates [Internet]. Available: <https://www.academy-info.go.kr/uipnh/unt/unmcom/RdViewer.do>.
- [5] M. H. Chung and Y. J. Yang, “A study on basic learning ability support system for university students: based on professors and students' perception and needs,” *JOEC*, vol. 22, no. 2, pp. 101-126, 2016.
- [6] J. G. Hwang, “Development and validation of the learning strategy scale for college students,” *Korean Journal of Counseling*, vol. 12, no. 5, pp. 1833-1855, 2011.
- [7] S. H. Kim, “Constructivist approach to improving practical vocational education at Korean junior colleges,” *Andragogy Today*, vol. 3, no. 3, pp. 99-126, 2000.
- [8] J. W. Hwang, H. J. Kim, and O. S. Song, “The relationship between course grades and the K-CESA core competencies for engineering students,” *Journal of Engineering Education Research*, vol. 19, no. 4, pp. 35-46, 2016.
- [9] H. Y. Kim and S. J. Lee, “Application plans of competence assessment: Focusing on the connection between assessment and curriculum,” *Korean Journal of General Education*, vol. 7, no. 4, pp. 139-172, 2013.
- [10] I. J. Ju, D. K. Kim, J. T. Jung, and H. H. Kim, “The analysis on the actual condition of development of competency model and application in corporation,” *The Journal of Vocational Education Research*, vol. 29, no. 3, pp. 309-334, 2010.
- [11] A. H. Lee and M. S. Choi, “Analysis of the research trend on the college students key competencies and diagnosis tools,” *Journal of Educational Technology*, vol. 30, no. 4, pp. 561-588, 2014.
- [12] J. D. Klein, J. M. Spector, B. L. Grabowski, and I. De la Teja, “Instructor competencies: standards for face-to-face, online, and blended settings,” Information Age Publishing, October 2004.
- [13] Center for Teaching and Learning, “Core Competence Center Curriculum Development Manual,” Doowon Technical University, pp. 1-35, 2018.
- [14] K. H. Jung, “Development research of specialized high school employment prediction model utilizing machine learning,” KERIS, 2020.



김 성 국 (Sungkook Kim)_정회원

1998년 2월 : 한국항공대학교 항공통신공학과 졸업(공학사)
 2005년 2월 : 한양대학교 컴퓨터공학과 졸업(공학석사)
 2018년 9월 ~ 현재 : 코리아텍 전기전자통신공학부 박사과정
 1999년 4월 ~ 2016년 8월 : 한국과학창의재단 선임연구원
 2016년 9월 ~ 현재 : 두원공과대학교 IT융합학부 교수
 <관심분야> 컴퓨터공학, 데이터베이스, 빅데이터, 인공지능, 머신러닝, 교육학 등



오 창 현 (Chang-Heon Oh)_종신회원

1988년 2월 : 한국항공대학교 항공통신공학과 졸업(공학사)
 1990년 2월 : 한국항공대학교 항공통신공학과 졸업(공학석사)
 1996년 2월 : 한국항공대학교 항공전자공학과 졸업(공학박사)
 1990년 2월 ~ 1993년 8월 : 한진전자(주) 기술연구소 전임연구원
 1993년 10월 ~ 1999년 2월 : 삼성전자(주) CDMA 개발팀 선임연구원
 1999년 2월 ~ 현재 : 코리아텍 전기전자통신공학부 교수
 <관심분야> 이동통신, 무선통신, Wireless Sensor N/W, 실천공학교수법, 산학협력교육 등