

# 빅데이터를 활용한 핵심역량과 학습역량과의 연계성 분석

## - D대학 사례를 중심으로 -

김성국<sup>1</sup> · 오창현<sup>2,\*</sup>

<sup>1</sup>두원공과대학교 · <sup>2</sup>한국기술교육대학교

## Analysis of core competence and learning ability using Big Data

### - Focusing on the case of D-University -

Sung-kook Kim<sup>1</sup> · Chang-heon Oh<sup>2,\*</sup>

<sup>1</sup>Doowon Technical University · <sup>2</sup>Korea University of Technology and Education

E-mail : sayes@doowon.ac.kr

### 요 약

핵심역량은 개인이 지속적으로 개발해야 하는 역량으로 학교 교육과정에 반영되어야 하는 개념으로 중요성이 더 확대되고 있다. 본 연구에서는 국내 D대학 졸업자를 대상으로 핵심역량중심 교육과정을 이수한 학생들의 핵심역량과 졸업학점(학습역량)의 관계를 분석하여 D대학의 교육과정 반영과 대학교육 정책을 개선하는 것을 목적으로 한다. 향후 분석 결과를 활용하여 D대학의 인재상에 부합하는 인재육성을 위한 자료로 활용할 계획이다.

### ABSTRACT

The core competence is further growing in importance as a concept that must be continuously developed by individuals and should be reflected in the school curriculum. In this study, we analyze the relationship between core competence and graduation credits (learning ability) of students who have completed a core competence-centered curriculum for graduates of D-University, and reflect it in the curriculum of D-University and university education policy. The purpose is to improve. We plan to utilize the results of future analysis as materials for human resource development that matches the human resources image of D-University.

### 키워드

Big Data, Core competency, curriculum, learning ability

### I. 서 론

본 연구의 목적은 D대학 졸업생의 핵심 역량 진단 결과를 바탕으로 빅데이터 분석을 통해 학습역량과의 관련성을 예측해 보는 것이다. D대학 졸업생의 핵심 역량을 확인하기 위해 핵심 역량 측정 도구를 사용하여 계열별, 개인별로 분석하였다. D대학의 핵심 역량 측정 방법은 5대 대학의 인재상과 연계된 10대 핵심역량을 2~3개 내외로 개별 교과목에 포함시켜 교육과정을 개발·운영하여 성적

을 산출한 후 이를 학기 단위로 측정하는 방식을 사용하고 있다. 이런 과정을 통해 획득한 계열별, 개인별 취업 가능성을 분석하고 이를 차년도 교육과정과 연계하여 취업률을 향상시키기 위한 방법으로 향후 추가 연구를 진행할 예정이다.

### II. 핵심역량에 따른 학습역량 분석 모델 선정

핵심역량 교육이 적용된 2017년 이후 졸업생을 대상으로 여러 가지 분석기법을 활용하여 학습역량의 우수 여부를 예측하는 실험을 실시하였다. 우

\* corresponding author

선 학습역량의 우수 여부는 대학에서 일반적으로 우수학점으로 생각하는 A학점(100점 환산으로 90점) 이상으로 하여 90점 이상일 경우 “T”, 89점 이하일 경우 “F”로 목표변수를 설정하였다. 또한 분석 도구로는 슬로베이나 Ljubljana 대학에서 개발한 Orange SW를 활용하여 진행하였다.

우선 예측 시스템에서 가장 많이 활용되는 Random Forest, 신경망, 로지스틱 회귀분석, 앙상블 모형중 가장 많이 활용되는 ‘Adaboost’ 기법을 활용하여 생성된 모델로 평가를 해보았다.

표 1. Test & Score

모델 \ 값	AUC	CA	F1	Precision	Recall
<b>Random Forest</b>	<b>0.947</b>	<b>0.908</b>	<b>0.906</b>	<b>0.906</b>	<b>0.908</b>
신경망	0.951	0.905	0.904	0.903	0.905
로지스틱 회귀분석	0.948	0.894	0.892	0.891	0.894
AdaBoost(앙상블)	0.846	0.871	0.869	0.868	0.871

[표 1]은 4개 분석 모델에 대한 ‘Test & Score’로 모델에 따른 성능을 보여주고 있다. AUC, CA (분류정확도), F1, Precision(정밀도), Recall(재현률) 모두 1에 가까울수록 좋은 모델로 판단할 수 있다. 결과에서 보듯 “Random Forest” 기법이 가장 좋은 모델로 판단되어 이 기법에 대한 Confusion Matrix 평가도 실시하여 예측모델에 대한 오류를 최소화하였다.

		Predicted		Σ
		N	Y	
Actual	N	95.3 %	4.7 %	4061
	Y	24.5 %	75.5 %	1217
Σ		4170	1108	5278

그림 1. Random Forest 기법에 대한 Confusion Matrix

[그림 1]을 보면 학습역량 우수자(Y)를 우수하다고 예측한 확률은 75.5%이고 학습역량 비우수자(N)를 우수하지 않다고 예측한 확률은 95.3%로 예측의 정확도가 준수한 것으로 판단되었다. 다만 학습역량 우수자를 비우수자로 잘못 예측한 확률이 24.5%로 다소 높게 나타나고 있다.

### III. 계열별 핵심역량에 따른 학습역량 분석

표본의 수가 부족한 경우 예측의 정확도가 떨어지므로 학과별 분석은 계열별 분석으로 대체하

여 진행하였다. 또한 모든 계열에 대해 상관분석을 시행하였으며 학습역량(졸업성적)에 어떤 역량이 주요하게 작용하는지 알아보았다. 이 분석에서는 학기별 역량이 아닌 모든 학기의 역량별 성적의 평균을 사용했으며, 어떤 역량이 전공 연계 여부에 핵심적인지 파악할 수 있었다.

표 2. 계열별 Test & Score

계열(학과수)	AUC	CA	F1	Precision	Recall
간호보건계열(5)	0.963	0.901	0.900	0.899	0.901
공학계열(14)	0.928	0.900	0.896	0.895	0.900
방송디자인계열(8)	0.893	0.877	0.875	0.874	0.877
인문사회계열(11)	0.955	0.906	0.906	0.905	0.906

계열별 정확도는 간호보건계열이 0.963으로 가장 높게 나타났으며 방송디자인계열이 0.893으로 가장 낮게 나타나고 있다. 이는 표본수의 차이가 가장 큰 요인으로 표본수가 많을수록 높은 정확도를 보이고 있는 것으로 나타나고 있다. 또한 계열내 속한 학과가 적을수록 예측의 정확도가 높게 나타나고 있다. 이는 학습역량은 학과별로 여러 가지 복합적인 요인이 발생하고 있어 학과의 다양성이 적을수록 높은 정확도를 나타내고 있는 것으로 판단된다.

		N	Y	Σ			N	Y	Σ	
N	94.3 %	5.7 %	474	Y	20.5 %	79.5 %	190	1458		
	Σ	486	178		664	Σ	1522	273	1795	
					[간호보건계열]					
N	93.7 %	6.3 %	303	Y	30.9 %	69.1 %	97	621		
	Σ	314	86		400	Σ	633	201	834	
					[공학계열]					
N	94.7 %	5.3 %	621	Y	21.1 %	78.9 %	213	834		
	Σ	633	201		834	Σ	633	201	834	
					[방송디자인계열]					
					[인문사회계열]					

그림 2. 계열별 Confusion Matrix

[그림 2]를 보면 계열별 예측의 정확도는 학습역량 비우수자(N)를 비우수자로 예측한 확률은 93.7%~96%로 모든 계열에서 높게 나타나고 있으며 학습역량 비우수자를 우수자로 잘못 예측한 확률은 4%~6.3%로 매우 준수한 예측률을 보이고 있는 것으로 나타나고 있다. 다만 학습역량 우수자(Y)를 비우수자로 잘못 판단할 확률이 계열에 따라 20.5%~36.2%로 높게 나타나 이에 대한 보완이 필요할 것으로 보인다.

IV. 고찰 및 향후 연구

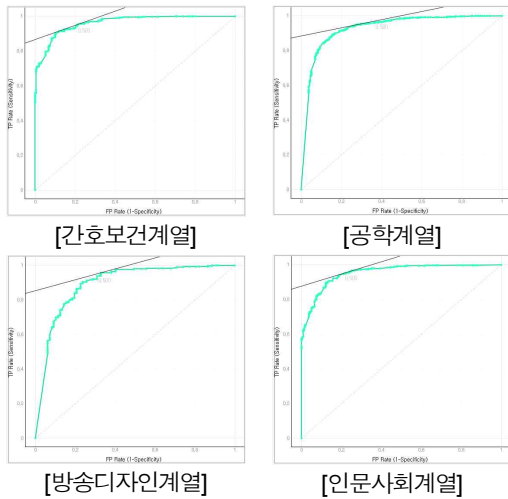


그림 3. 계열별 ROC curve

[그림 3]의 ROC (Receiver Operating Characteristic) curve는 민감도를 분석하는 자료로 여러 임계값들을 기준으로 Recall-Fallout의 변화를 시각화한 자료이다. X축은 실제 “F”인 데이터 중에서 모델이 “T”로 분류한 비율이고 Y축은 실제 “T”인 데이터 중에서 모델이 “F”로 잘못 분류한 비율을 나타내고 있다. 간호보건계열이 좌상단에 근접하므로 가장 우수한 모델이고 방송디자인계열은 완만하게 수렴하므로 가장 부적절한 모델로 나타나고 있다. 이는 앞에 기술된 바와 같이 표본의 수가 적고 계열내 학과의 수가 많아 다양성을 모두 수렴하지 못한 결과로 보인다. 이를 위해 향후 지속적인 교육과정의 데이터화를 통해 표본의 수를 늘려 나간다면 우수한 예측모델로 손색이 없을 것으로 보인다.

본 연구는 국내 D대학의 졸업생을 대상으로 핵심역량기반 교육과정을 수강한 학생들의 핵심역량과 학습성과(졸업성적)와의 관련성을 분석하여 D대학 교육과정의 개선 및 대학 교육 정책에 반영하는 것을 목적으로 한다. 향후 분석 결과를 활용하여 학습역량 강화 및 D대학의 인재상에 맞는 인재를 육성하기 위한 자료로 사용할 예정이다.

References

[1] Seon-Ho Jeong, “Analyzing Effect of General Education based on Core competences” *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, Vol. 6, No. 12, pp. 253-262, December. 2016.

[2] Hwang, KyungSoo, Kwon, SoonChul, Koh, BongJo, “Study on Analysis of the core competencies collegiate students recognized” *Korean Journal of General Education*, Vol. 9, No. 3, pp. 131-162, September. 2015.

[3] Jang Ik Lee, “A Study on University Students' Essential Skills and University Admission Methods” *Journal of Agricultural Education and Human Resource Development*, Vol. 44, No. 2, pp. 73-96, June. 2012.

[4] Yeong-Ju Kwo, “A Review and Prospect of Korean University Student’s Carer Competencies” *The Journal of Employment and Carer*, Vol. 3, No. 2, pp. 21-39, December. 2013.

[5] Kim, Sunny, Kwon, Youngshim, Shin, Jisook, “A study for Promoting Key Competencies of College Students” *Korean Corporation Management Review*, Vol. 19, No. 2, pp. 57-77, December. 2012.