Investigating Predictors of First Graders' School Adjustment in Korean Elementary School Using Latent Profile Analysis and Machine Learning

Eun Kyoung Goh¹, Hyo Jeong Jeon²

Assistant Professor, Human Life Research, Dong-A University, Busan, Korea¹

Professor, Department of Child Studies, Dong-A University, Busan, Korea²

잠재프로파일분석과 머신러닝을 활용한 한국 초등학교 1학년 아동의 학교적응 예측

고은경¹, 전효정²

동아대학교 휴먼라이프리서치센터 조교수¹, 동아대학교 아동학과 교수²

Objectives: The purpose of this study is to evaluate the factors of first graders' school adjustment in Korean elementary school by using latent profile analysis and machine learning to predict and classify among the first-graders who have difficulty in school adjustment.

Methods: Data from 1,010 children in the first grade of elementary school were used in the seventh to eighth surveys of the 2014-2015 Korean Children's Panel. The output variable of machine learning is the school adjustment type of the eighth survey data, and the type and number of each school adjustment were determined by latent profile analysis of its sub-variables. The input variables for machine learning were the children's gender and school readiness of the seventh data. In addition, the children's difficulty in executive function, school preference, parents' interest in school life, teacher's teaching efficacy, and work stress included in the eighth data were selected as input variables. As a model for predicting children's school adjustment type, four machine learning algorithms are used.

Results: Three types of school adjustment were identified through latent profile analysis, with the lowest level of school adjustment being 13.1% of all children. "Difficulty in executive function" was the predictor of the lowest level of school adjustment in all machine learning models. The model that best classifies children at the lowest level of school adjustment was the gradient-boosted decision tree model, with an accuracy of 94%, a sensitivity of 70%, a specificity of 97%, and an AUC of .89.

Conclusion: When using latent profile analysis and machine learning model, the difficulty in adjusting to school in the first graders of elementary school was most predicted by difficulty of the executive function; however, school readiness was not. This study suggested that cooperation between early childhood education and primary education based on empirical analysis is necessary for successful primary connection.

Keywords: Korean elementary school, first graders' school adjustment, latent profile analysis, machine learning, discriminative model

Introduction

초등학교 입학 후 첫 1년은 아동이 본격적으로 학령기에 진입하는 시기로서 12년 동안 지속되는 우리나라 초·중등교육체제의 시작 단계이다. 초등학교 입학 전후는 발달적으로 유사한 시기임에도 불구하고 교육과정의 성격은 질적으로 많이 달라지는데, 5세 누리과정은 일상과 놀이로 자연스럽게 통합되는 성격을 갖지만, 초등학교 1학년 교육과정은 국어, 수학 등의 교과 중심의 교육과정이 특징이다(Hwang, 2012). 교실에서의 활동 방식도 유아교실에서는 충분한 자유선택활동 시간과 아동이 자유로운 데 반해, 초등학교 1학년 교실에서는 40분 동안의 수업시간 동안 착석을 유지하는 것이 일반적이다. 유아기와 달리 사회적 규범, 학업적 수행, 대인관계에 있어 이전보다 높은 수준의 수행을 기대받기 때문에 이러한 요구에 익숙하지 않은 아동은 쉽게 스트레스를 경험하게 된다(Jeong & Park, 2017; M. Kim, 2008). 아동과 부모의 관점에서 이 시기의 스트레스를 극복하고 안정적으로 발달하는 것은 아동의 정신건강과 행복한 삶을 위해 매우 중요한 과업일 수밖에 없으며, 사회적인 차원에서도 적극적으로 지원할 필요가 있다. 더욱이 초등학교에서의 학교적응이 이후 중고등학교까지 학업적인 성취와 학교생활에 지속해서 영향을 미친다(M. Kim, 2008; Simons-Morton & Crump, 2003)는 사실을 고려할 때, 아동의 학교적응을 지원하는 것은 우리 사회 구성원을 건강하게 성장시키는 가장 기본적인 작업이다.

인간에게 적응은 일상생활의 문제나 요구에 대처하거나 관리하는 심리적 과정으로서(Weiten, Lloyd, Dunn, & Hammer, 2009), 학교적응은 학습자인 아동이 학교에서 요구하는 과제와 능력의 목표를 이해하고 성취하며, 학교생활과 인간관계가 요구하는 규칙을 준수하고 갈등에 대처하면서 심리적으로 성숙하는 과정이다(Roeser & Eccles, 1998). 그런데 초등학교 입학 직후의 학교적응은 유아기와는 이질적인 교실문화에 적응해야 한다는 점에서 아동에게 도전적인 과업이 된다. 초등학교사의 엄격한 태도, 규격화된 교실 구조와 일과, 지정된 좌석으로 인한 움직임의 제한과 그로 인한 또래 관계의 제한, 과제수행의 어려움이 증가하는 것은 아동들에게 문화적인 충격이 된다(Chi & Jung, 2006). 그러나 이러한 문화적 충격을 극복하면서 점차 아동은 학교생활에 적응하게 되는데, 한국아동패널의 종단자료를 활용하여 초등학교 1학년부터 3학년까지 측정한 학교적응은 해마다 향상되었다(K. Lee, 2020). 반면, 초등학교 고학년 이후부터는 학년이 올라갈수록 학교적응 수준은 감소하게 된다. 즉, 초등학교 입학 후 3년 동안 1학년의 학교

적응 수준이 가장 낮지만(J. Lee, Kang, Lee, & Yi, 2015; K. Lee, 2017) 고학년보다는 전반적으로 높다(K. Lee, 2020). 이러한 차이의 원인으로는 저학년 시기의 낮은 학업성취 압력과 환경적인 지원을 들 수 있다. 우리나라 아동에게 학업성취 압력은 초등학교 저학년부터 학교적응에 분명 영향을 미치지만(K. Lee & Lee, 2013), 입학 후 2학년까지는 개별 교과목 중심의 학습이 아닌 통합적인 교육과정이 운영되어 학업에 대한 부담감이(M. Sung, Chang, & Seo, 2016)과 학업성취 압력(Hong & Lee, 2012)이 고학년과 비교해 낮다. 또한, 입학 직전 유아교사와 부모는 아동에게 초등학교에 대해 기대와 관심을 고무시키는데 이러한 격려가 초등학교 입학 직후의 학교적응에 도움을 줄 수 있다(Jeong & Park, 2017). 따라서 일반적으로 초등학교 1학년의 학교적응은 고학년과 비교해 높은 수준이지만 교실문화의 이질성으로 인해 적응상의 위기를 경험할 수 있다.

그러나 이러한 일반적인 학교적응의 양상도 개인이나 집단별로 달라질 수 있다. 개별 변수들의 관계에 초점을 두는 변수중심적 접근(variable-oriented approach)과 달리 사람중심적 접근(person-oriented approach)은 특정 모수치에 따라 이질적인 집단으로 분류하고 유형화하는데 초점을 둔다(Berman & Magnusson, 1997). 이러한 사람중심적 접근의 분석은 개인이나 집단의 특성을 지원하는데 실증적 근거가 될 수 있다. 우리나라에서는 한국아동패널, 아동·청소년패널과 같은 대규모 2차 자료를 활용하게 되면서 이질적인 발달특성을 보이는 사람중심적 접근의 연구가 증가하였다(Y. H. Kwon, Jeon, Goh, 2019). 대표적으로 잠재프로파일분석은 이질적인 발달 양상을 집단별로 규명하는 대표적인 사람중심적 접근의 분석방법 중 하나이다. 이러한 잠재프로파일분석을 적용하여 학교적응에 어려움을 보이거나 특수한 지원이 필요한 아동을 선별할 수 있다. 우리나라 학교적응에 대한 잠재프로파일분석 연구로 초등학교에서 중학교로의 학교적응 전이를 분석한 연구(Park, 2019)에 의하면 중학교 전이 과정에서 30%에 달하는 학생들이 부적응을 경험하며 여학생과 자아존중감이 낮을수록 부적응의 위험이 컸다. 그러나 유아기에서 학령기로의 전이 양상은 발달과 환경적 차이로 인해 중학교로의 전이 양상과 다를 수밖에 없으며, 따라서 실증적인 근거에 기반을 두어 초등학교 적응을 지원하기 위해서는 이 시기에 대한 잠재프로파일분석과 예측요인을 도출하는 분석이 필요하다.

생태학적 체계이론(Bronfenbrenner, 1979)에서 학교는 가족과 더불어 아동의 대표적인 미시체계이다(W. Kim & Kwon, 2019; Park, 2019). 아동의 생태체계를 구성하는 요소 중에서 아동과 미시체계는 가장 직접적이며 상호적으로 영향을 미치

는 요소이다. 따라서 학교적응의 예측요인으로 아동의 개인적 특성, 가족의 특성, 학교의 특성을 가장 우선으로 검토할 필요가 있다. 아동의 학교적응을 다면적으로 분석한 선행연구들에서도 예측요인을 크게 아동의 개인적 요인, 부모와 가족 요인, 교사 및 학교 요인으로 구분하고 있다(C.-I. Kim, 2018; E. Kim, 2018; J. Kim & Hong, 2019a; K. Lee & Lee, 2013). 먼저, 아동 개인적 요인으로는 성별, 집행기능과 같은 심리적 요인, 초등학교 전이를 위한 학교준비도 등을 들 수 있다. 성별이 학교적응에 미치는 영향에 대해서는 일반적으로 남아보다 여아의 초기 학교적응력이 높은 것으로 알려져 있으나(C.-I. Kim, 2018; E. Kim, 2018; J. Kim & Hong, 2019a; Roeser & Eccles, 1998; Simons-Morton & Crump, 2003) 차이가 없다는 보고도 있어(M. Y. Jung & Moon, 2007) 보다 명확한 규명이 필요하다. 아동의 학교적응에 영향을 주는 심리적 요인 중에서 대표적으로 집행기능은 목적달성을 위해 자신의 행동과 정서적 반응을 조절하는 상위인지적 기능(Barkley & Lombroso, 2000)으로서 초등학교 입학 전후로 5-8세에 급격히 발달한다(Korkman, Kemp, & Kirk, 2001). 한국의 초등학교 1학년 아동의 집행기능 곤란은 학교적응과 부적 관계였으며(S. Kim, 2018), 회귀분석에서도 미국의 노숙 가정의 유아와 초등학교 1학년 아동(Masten et al., 2012), 한국의 초등학교 1학년 아동(Yi & Jun, 2019)을 대상으로 한 선행연구에서 집행기능 곤란이 낮을수록 높은 학교적응 수준을 예측하였다. 또한, 집행기능 곤란은 학업성취에도 장기적으로 영향을 끼치는데, 미국 교육부의 코호트 자료를 활용하여 유치원에서 초등학교 3학년에 걸쳐 읽기, 수학, 과학의 학업성취 궤적을 추적한 종단연구에서 학업성취 수준이 낮은 집단일수록 집행기능 곤란이 낮았다(Morgan, Farkas, Hillemeier, Pun, & Maczuga, 2019). 이처럼 집행기능 곤란은 초등학교 저학년의 학교적응 수준을 예측하는 주요 요인이 될 수 있음에도 그동안 학교적응을 다차원적으로 예측한 연구(E. Kim, 2018; J. Kim & Hong, 2019a; W. Kim & Kwon, 2019)에서는 간과되었다.

한편, 학교준비도는 학령기 학교생활 준비를 위한 인지적 발달에서 사회·정서적, 신체적 발달까지 아우르는 다차원적이고 포괄적인 상태라고 할 수 있다(Kagan & Kauerz, 2007). 이러한 학교준비도는 아동의 학교적응에 긍정적으로 영향을 미치는 주요 변수였다(Y. A. Jung & Kim, 2019; J. Kim, 2019; J. Kim & Hong, 2019b; Y. H. Kim, Kim, & Kim, 2019; S. A. Lee & Lim, 2019). 그뿐만 아니라, 라틴계 아동의 종단자료를 활용한 연구에서 높은 수준의 학교준비도로 입학한 아동집단이 초등학교 2학년부터 5학년에 걸쳐 읽기 성취가 높은 집단으로 전이되

는 것을 예측하였다(Quirk, Grimm, Furlong, Nylund-Gibson, & Swami, 2016). 학교준비도는 초등학교 생활을 대비하는 대표적인 변인으로서 유아기 아동발달의 다양한 측면을 포괄하고 있다. 그러나 본격적인 학업 수행이 요구되는 학령기에는 집중력과 조절력을 유아기보다 더 많이 요구할 수 있다. 따라서 아동의 집행능력과 학교준비도가 학교적응에 어느 정도 영향을 미치며 학교적응 수준을 집단별로 구분하는 데 어느 정도 기여하는지 비교할 필요가 있다.

다음으로, 부모와 가족 요인 중 자녀의 학교적응 예측요인으로 부모의 자녀 학교생활 지원을 들 수 있다. 부모가 자녀의 학교생활이나 학업적 성취에 관심을 가지고 자녀와 의사소통하는 것은 자녀의 학교적응을 돕는 대표적인 부모 지원활동이다(Jervis & Pollard, 2007, as cited in J.-H. Lee & Kim, 2020). 부모가 가정에서 자녀의 학업적 성공을 위해 직접 지도하거나 학습 분위기를 조성하는 것은 사교육에 의존하거나 자녀의 학습을 통제하는 것보다 아동의 학교적응에 더 긍정적이었다(No, Park, Park, & Choi, 2020). 나아가 부모가 학교와 상호작용하는 학교참여(Hill & Tyson, 2009)도 자녀의 학교적응에 기여하였다(J. Y. Kim & Lee, 2020; Wyrick & Rduasill, 2009, as cited in E. Kim, 2018). 부모의 자녀 학교생활 관심도는 자녀의 학교생활에 대해 부모가 전반적으로 관심을 가지는 정도를 교사가 평가하는 것으로(E. Kim, 2018), 사실상 부모의 학교참여를 통해 평가하는 것이다. 실제로 부모의 자녀 학교생활 관심도가 높을수록 1학년 아동의 학교적응에 긍정적인 영향을 미쳤는데(E. Kim, 2018; J.-H. Lee & Kim, 2020), 이러한 부모의 관심이 학교적응에 어려움을 겪는 아동을 예측하고 분류하는데 어느 정도 기여하는지 알아볼 필요가 있다.

마지막으로 학교적응에 영향을 미치는 학교 요인으로는 아동의 학교에 대한 평가와 교사 관련 요인으로 구분할 수 있다. 긍정적인 학교 환경은 아동의 학교생활에 긍정적인 영향을 줄 수 있지만, 특히 아동이 학교에 대해 주관적으로 평가하는 학교선호도는 학교적응에 직접적인 관련성을 가질 수 있다. 아동의 학교선호도는 아동이 학교와 학교생활을 어느 정도로 좋아하는지 평가하는 것으로, 아동의 학교선호도가 높을수록 학교적응 수준도 높았다(W. Kim & Kwon, 2019). 다음으로, 교사 효능감은 학생의 학업 및 행동지도에 미치는 교사의 긍정적인 영향(K.-S. Sung, Kim, Park, & Min, 2012)을, 교사의 업무스트레스는 교사의 직무영역에서 경험하는 스트레스 정도(J. Kim & Hong, 2019a)를 스스로 평가하는 것이다. 교사의 효능감이 높을수록 아동의 학교적응 수준은 높아지지만, 교사의 업무스트레스가 높을수록 아동의 학교적응 수준은 낮았다(J. Kim &

Hong, 2019a). 따라서 아동을 지원하기 위한 실증적 근거로서 학교적응을 예측하는 다양한 요인과 함께 교사효능감 및 업무 스트레스와 같은 교사요인이 어느 정도로 기여하는지 비교할 필요가 있다.

선행연구들에서 학교적응의 영향요인을 분석하는 방법은 주로 회귀분석 방식에 의존해왔다. 그러나 머신러닝 기법을 적용하면 학교적응에 대한 예측뿐만 아니라 학교적응에 성공적이거나 어려움을 겪는 아동을 분별할 수 있다. 머신러닝은 크게 과거 데이터를 학습하여 미래를 예측·분류하는 지도학습(supervised learning)과 데이터의 숨겨진 패턴이나 구조체를 발견하는 비지도학습(unsupervised learning)으로 구분할 수 있다. 지도학습에 의한 분별모델 알고리즘에는 로지스틱 회귀, 라쏘 로지스틱 회귀, 랜덤포레스트, 그레디언트-부스티드 의사결정나무 모델이 있다. 이러한 지도학습의 머신러닝 모델을 한국아동패널자료의 학교적응에 적용하면, 전체 아동을 위한 보편적인 지원뿐만 아니라 개별 아동에 대한 분류와 지원을 위한 근거로도 활용할 수 있다.

본 연구의 목적은 한국아동패널 7·8차 자료 중에서 초등학교 1학년 아동의 학교적응을 예측하고 분류하기 위해 잠재프로파일분석으로 학교적응의 집단유형을 규명하고, 로지스틱 회귀, 라쏘 로지스틱 회귀, 랜덤포레스트, 그레디언트-부스티드 의사결정나무의 머신러닝 모델을 학교적응 집단유형을 예측하는 모델로 적용하여 모델별로 예측력과 분류 성능을 비교한 후, 최종적으로 학교적응을 예측하는 변수들의 상대적 기여도를 확인하고자 한다. 이러한 일련의 분석은 아동의 학교적응을 사람중심적으로 지원하는 실증적인 근거 자료가 될 것이다. 본 연구의 문제는 아래와 같다.

연구문제 1

초등학교 1학년 아동의 학교적응 유형은 어떠한가?

연구문제 2

머신러닝 모델별로 아동의 학교적응 유형에 대한 예측력은 어떠한가?

연구문제 3

머신러닝 모델별로 아동의 학교적응 유형을 예측하는 주요 변수는 무엇인가?

Methods

연구대상

본 연구의 대상은 2008년에 출생하여 2015년에 초등학교 1학년이 된 아동이다. 분석자료는 육아정책연구소에서 2014년에 실시된 한국아동패널 7차 조사 자료와 2015년에 실시된 한국아동패널 8차 자료를 활용하였다(Panel Study on Korean Children [PSKC], 2014, 2015). 7차의 학교준비도와 8차의 변수들을 분석에 활용하기 위하여 두 시기의 자료를 모두 활용할 수 있는 1,010명의 자료를 최종적으로 선정하였다. 본 연구대상 아동의 성별은 여아가 492명, 남아가 518명으로 여아가 전체의 48.7%로 남아보다 적었다.

연구도구

아동의 학교적응

아동의 학교적응 자료는 2015년 실시된 한국아동패널 8차 자료이며, 아동이 학교생활에서 요구하는 4가지 영역, 즉 질서와 규칙과 관련된 학교생활, 학업수행, 또래관계, 교사와의 관계에서 적응적인지 평가한 것이다(Chi & Jung, 2006). 총 35개의 문항에 대한 응답은 아동의 교사가 수행하였으며, 5점 Likert 척도로 점수가 높을수록 아동의 학교적응이 높음을 의미한다. 본 연구의 대상인 1,010명의 자료에서 학교적응의 신뢰도(Cronbach's α)는 .95이다.

아동의 학교준비도

아동의 학교준비도는 2014년 실시된 7차 자료이며, 초등학교 생활을 위해 학습에 대한 태도, 사회정서발달, 인지발달 및 일반적 지식, 의사소통 영역에서 유아가 발달적으로 준비된 상태를 평가한 것이다(Murphey & Burns, 2002). 총 22문항에 대한 응답은 유아교사에 의해 이루어졌으며, 4점 Likert로 점수가 높을수록 아동의 학교준비도가 높음을 의미한다. 본 연구의 대상인 1,010명의 자료에서 학교준비도의 신뢰도(Cronbach's α)는 .93이다.

아동의 집행기능 곤란

아동의 집행기능 곤란은 2015년 실시된 8차 자료이며, 아동

Table 1

Descriptive Statistics of Children

Variables		<i>M</i>	<i>SD</i>	Min	Max	<i>n</i>	%
Gender	Boys					518	51.3
	Girls					492	48.7
School adjustment	School life	4.08	0.94	1.09	5.00		
	Academic performance	3.83	0.86	1.00	5.00		
	Peer	4.09	0.81	1.00	5.00		
	Teacher	4.21	0.69	2.40	5.00		
School readiness		3.55	0.36	2.00	4.00		
Difficulty in executive function		1.31	0.40	1.00	3.00		
School preference		3.24	0.87	1.00	4.00		
Parent's interest in school life		4.12	0.84	1.00	5.00		
Teacher's teaching efficacy		4.34	0.47	2.11	5.00		
Teacher's work stress		3.06	0.84	1.00	5.00		

Note. *N* = 1,010.

이 목적 수행을 위해 자기조절이 필요한 영역, 즉 주의, 정서통제, 행동통제, 계획-조직화에서 곤란한 정도를 평가한 것이다 (Song, 2014). 총 40문항에 대해 교사가 응답한 자료를 이용하였으며, 3점 Likert 척도로 점수가 높을수록 집행기능에 어려움이 있는 것이다. 본 연구대상 1,010명의 자료에서 집행기능 곤란의 신뢰도(Cronbach's α)는 .97이다.

부모의 자녀 학교생활 관심도

부모의 자녀 학교생활 관심도는 한국아동패널 8차 자료이며, 부모가 자녀의 학교생활에 관심이 있는 정도를 묻는 1개의 문항에 대해 교사가 평가한 것이다. 5점 Likert로 점수가 높을수록 부모가 자녀의 학교생활에 관해 관심의 정도가 높음을 의미한다.

아동의 학교선호도

아동의 학교선호도는 한국아동패널 8차 자료의 단일 문항을 이용하였으며, 아동이 학교와 학교생활을 어느 정도로 좋아하는지 평가한 것이다. 아동에게 학교에 대해 좋아하지 않는다(1점)에서 매우 좋아한다(4점)까지 4점 Likert로 응답하게 되어 있다. 점수가 높을수록 학교선호도가 높음을 의미한다.

교사효능감

교사효능감은 한국아동패널 8차 자료이며, 학생의 성취에 영향을 미치는 교사 자신의 효능감에 대해 스스로 평가하는 것

이다. 도구는 K.-S. Sung 등(2012)이 Im 등(2010)의 문항을 수정한 것이다. 학생의 학업, 생활지도, 행동지도 등과 관련된 총 9문항에 대해 5점 Likert 척도에 응답하게 되어 있다. 점수가 높을수록 교사의 효능감이 높음을 의미한다. 본 연구자료에서 교사효능감 신뢰도(Cronbach's α)는 .90이다.

교사의 업무스트레스

교사의 업무스트레스는 한국아동패널 8차 자료이며, 학부모 관계, 생활지도, 수업지도, 행정업무에 대한 스트레스를 묻는 4개의 문항에 교사가 응답하게 되어 있다. 척도는 전혀 스트레스를 받지 않는다(1점)에서 스트레스를 많이 받는다(5점)까지 5점 Likert 척도에 응답하게 되어 있다. 점수가 높을수록 교사의 업무스트레스가 높음을 의미한다. 본 연구자료에서 교사 업무스트레스의 신뢰도(Cronbach's α)는 .72이다.

머신러닝 모델에 투입되는 예측변수들은 일반적으로 모델 성능을 향상시키거나 변수 간의 중요도를 상대적으로 평가하기 위하여 0-1 범위로 정규화하거나 평균을 0, 표준편차를 1로 표준화하여 모델에 투입한다. 본 연구에서는 선정된 예측변수를 표준화하여 투입하였다. 한국아동패널 7-8차 자료에서 연구대상의 예측변수 측정값을 표준화 이전의 값으로 제시하면 Table 1과 같다.

연구절차

본 연구의 연구절차는 3단계로 이루어졌다. 첫째, 학교적응 하

위변수에 대한 잠재프로파일분석을 수행하여 학교적응의 잠재집단 유형을 확인하였다. 둘째, 잠재집단 유형을 결과변수로 하는 머신러닝 모델을 개발하고 모델별로 성능평가를 실시하였다. 셋째, 머신러닝 모델 중에서 정확도가 높은 모델을 선택하여 잠재집단 유형별로 분류에 기여하는 예측변수의 중요도를 비교하였다.

자료분석

잠재프로파일분석

본 연구에서 아동의 학교적응 잠재집단을 도출하기 위해서 세 가지 준거를 사용하였다. 첫째, 정보적합도 중에서 아카이케 정보기준지수(Akaike Information Criterion [AIC]), 베이지언 정보기준지수(Bayesian Information Criteria [BIC]), 표본크기보정 베이지언 정보기준지수(Sample-Size Adjusted Bayesian Information Criteria [SABIC])를 이용하여 잠재계층의 수를 확인하였다. 이들 지수는 수치가 낮을수록 좋은 적합도를 나타낸다. 둘째, 부트스트래핑 우도비검정(Bootstrap Likelihood Ratio Test [BLRT])을 활용하여 잠재계층 성립의 유의성을 확인하였다. 셋째, 분류된 집단크기가 전체 표본에서 1.0% 이상이 되는 모델만을 비교하였다. 본 연구에서 잠재프로파일분석은 'tidyLPA' 패키지를 활용해 이루어졌다.

머신러닝 모델 구축 및 성능평가

머신러닝 모델 구축 및 성능평가 과정은 다음과 같다. 첫째, 머신러닝 모델의 결과변수와 예측변수를 선정하였다. 본 연구에서 결과변수는 학교적응에 대한 잠재프로파일분석 결과로 도출된 잠재집단 유형이다. 본 연구에서 잠재집단은 Low level, Middle level, High level의 3개의 집단으로 존재했기 때문에, 이를 구분하는 4개의 분별모델을 도출하였다. 첫 번째 분별모델은 전체 아동에서 학교적응의 Low level 집단을 1로, 나머지 집단을 0으로 구분하는 모델로 출현율(prevalence)은 13.1%이다. 두 번째 분별모델은 학교적응의 Low level 집단을 1로, High level 집단을 0으로 구분하는 모델로 출현율은 18.3%이다. 세 번째 분별모델은 학교적응의 Low level 집단을 1로, Middle level 집단을 0으로 구분하는 모델로 출현율은 31.3%이다. 마지막으로 네 번째 분별모델은 학교적응의 Middle level 집단을 1로, High level 집단을 0으로 구분하는 모델로 출현율은 33.0%이다. 이러한 잠재집단의 분류를 예측하는 변수(predictor)로는

아동의 성별, 유아기 학교준비도, 집행기능 곤란, 부모의 자녀 학교생활 관심도, 아동의 학교선호도, 교사의 효능감과 업무 스트레스를 선정하였다.

둘째, 머신러닝 모델은 로지스틱 회귀 모델을 참조모델로 하여, 라쏘 로지스틱 회귀, 랜덤 포레스트, 그레디언트-부스티드 의사결정나무의 4개 모델로 구축하였다. 머신러닝 모델이 훈련 데이터에서 과적합(overfitting)되면 새로운 데이터에서는 오류가 증가할 수 있어 정규화 혹은 일반화(regularization)가 필요하다. 이러한 오류를 제어하는 파라미터가 하이퍼-파라미터(hyper-parameter)이다. 일반적인 파라미터와 달리 하이퍼-파라미터는 모델 외부에서 분석자에 의해 설정된다(Kuhn & Johnson, 2013). 그러나 R 패키지를 활용하면 최적화된 하이퍼-파라미터를 확인할 수 있다. 우선, 라쏘(least absolute shrinkage and selection operator [lasso]) 회귀 모델은 로지스틱 회귀 모델에서 가중치의 절대값 합을 최소화시키는 방향으로 제약하는 모델이다. 이러한 최소가중치 람다를 통해 예측에 기여하지 않거나 상관성이 높은 변수의 계수를 0으로 바꾸어 모델을 단순화시키고 다중공선성 문제도 해소할 수 있다(Muthukrishnan & Rohini, 2016). 최소가중치 람다(minimal lambda)는 'glmnet' 패키지를 활용하여 10겹의 교차검증(10-fold cross-validation)으로 산출하였다. 다음으로, 랜덤 포레스트 모델은 부트스트래핑 방식으로 생성된 의사결정나무를 숲처럼 결합한 모델이다(Breiman, 2001). 이 모델에서 하이퍼-파라미터는 의사결정나무의 수와 깊이를 결정하는 것이다. 마지막으로, 그레디언트-부스티드 의사결정나무 모델은 의사결정 나무를 병렬적으로 생성시키는 것이 아니라 오차를 줄이면서 너무 깊지 않은 나무를 순차적으로 생성시키는 모델이다(Elith, Leathwick, & Hastie, 2008). 이 모델에서 하이퍼-파라미터는 적절한 학습률과 나무 수를 결정하는 것이다. 추가로 빠른 구동을 위해 'xgboost' 패키지를 활용하였다. 랜덤 포레스트 모델과 그레디언트-부스티드 의사결정나무 모델의 최적의 파라미터는 'caret' 패키지를 통해 10겹의 교차검증으로 확인하였다.

셋째, 머신러닝 모델의 성능평가는 교차검증을 통해 산출된 혼동행렬(Table 2)의 수치를 활용하여 정확도, 민감도, 특이도, AUC 수치를 산출하고, 추가로 ROC curve와 decision curve를 시각화하여 성능평가에 활용하였다. 일반적으로 머신러닝 모델의 성능을 평가하기 위한 교차검증은 데이터를 훈련용(train set)과 테스트용(test set)으로 나누어 실시한다. 본 연구에서는 연구대상 1,010명의 데이터에서 훈련용 데이터를 70%, 테스트용 데이터를 30%로 임의로 할당하여 모형화하였다.

Table 2*Confusion Matrix*

Reference class		Predicted class	
		Positive	Negative
	Positive	True positive	False negative
	Negative	False positive	True negative

Table 3*Latent Profile Analysis about School Adjustment*

Number of class	Log likelihood	SABIC	AIC	BIC	BLRT	Entropy	Latent class distribution rate (%)			
							1	2	3	4
1	-12794	25641	25617	25823		1.00	100.0			
2	-12645	25360	25327	25421	300**	.87	18.7	81.3		
3	-12587	25263	25221	25339	116**	.83	13.1	28.7	58.2	
4	-12587	25282	25231	25374	-0.13	.67	0.0	13.2	30.1	56.7

Note. $N = 1,010$. SABIC = sample-size adjusted bayesian information criteria; AIC = akaike information criteria; BIC = bayesian information criterion; BLRT = bootstrap likelihood ratio test.

** $p < .01$.

민감도(sensitivity)는 실제 양성인 데이터를 양성으로 예측한 비율(true positive rate)이고, 특이도(specificity)는 실제 음성인 데이터를 음성으로 예측하는 비율(true negative rate)이다. AUC 통계치는 민감도와 '1 - 특이도(false negative rate)'를 두 축으로 하는 ROC curve 아래 면적이다. AUC가 .7 이상이면 '허용되는(fair)' 수준, .8 이상이면 '양호한(good)' 수준, .9 이상이면 '우수한(excellent)' 수준으로 평가한다(Safari, Baratloo, Elfil, & Negida, 2016). Delong's test를 통해 모델 간 AUC 값의 차이를 검정할 수 있다(DeLong, DeLong, & Clarke-Pearson, 1988). 정확도, 민감도, 특이도는 'caret' 패키지를 활용하여 혼동행렬을 통해 산출하였다. AUC 값의 산출, Delong's test 실시, 그리고 ROC curve 시각화를 위해 'pROC' 패키지를 활용하였고, Vickers와 Elkin (2006)의 decision curve 분석을 위해 R code를 활용하였다.

머신러닝 모델 예측변수의 상대적 중요성

머신러닝 모델에서 예측변수의 상대적 중요성을 분석하기 위하여 참조모델인 로지스틱 회귀 모델의 회귀계수를 비교하였고, 성능평가를 통해 예측력이 뛰어난 머신러닝 모델 기법을 선정하여 예측변수의 상대적 중요도를 분석하였다. 먼저, 로지스틱 회귀모델에서 예측변수를 표준화 값으로 변환하여 회귀계수의 크기를 비교하였다. 다음으로, 머신러닝 모델 중에

서 예측 성능이 가장 뛰어난 기법의 모델을 선택한 후, 가장 중요한 변수의 중요도를 100으로 수치화하고 나머지 변수의 중요도를 비례적으로 적용하여 중요도를 비교하였다. 본 연구의 모든 분석은 R version 3.6.1에서 이루어졌다.

Results

학교적응 잠재프로파일 유형과 특성

학교적응 잠재프로파일 유형

본 연구대상의 학교적응 하위영역에 대한 잠재프로파일 유형은 Table 3과 같다.

Table 3에서, 2-3개의 잠재집단으로 분류할 때 BLRT 검정에서 귀무가설이 기각된다($p < .01$). 잠재계층분류율에서 모든 표본이 1.0% 이상의 비율을 보이는 모델도 2-3개의 잠재집단으로 분류할 때이다. Entropy도 2-3개의 잠재집단으로 분류할 때는 .8 이상이지만 4개의 잠재집단으로 분류할 때 .7 이하로 감소한다. 한편, AIC, BIC, SABIC의 값은 3개의 잠재집단으로 분류할 때 가장 값이 작다. 따라서 3개의 잠재집단을 갖는 모형이 최적의 잠재계층 분류모델로 선택된다.

3개의 잠재집단을 갖는 모델을 선택하여 학교적응의 잠재

Table 4

Comparison of Means by Latent Groups

Sub variable	Class	<i>n</i>	<i>M</i>	<i>SD</i>	<i>F</i>	Scheffé
School life	a	132	2.18	0.46	3518.81***	a < b < c
	b	290	3.61	0.36		
	c	588	4.73	0.28		
Academic performance	a	132	2.97	0.95	225.75***	a < b < c
	b	290	3.44	0.68		
	c	588	4.22	0.66		
Peer	a	132	3.02	0.98	333.13***	a < b < c
	b	290	3.54	0.56		
	c	588	4.36	0.55		
Teacher	a	132	3.72	0.80	26.35***	a = b < c
	b	290	3.86	0.68		
	c	588	4.14	0.75		

Note. *N* = 1,010. a = low level; b = middle level; c = high level.

****p* < .001.

Figure 1. The sub-variables of school adjustment by latent groups.

집단 유형별 특성을 표와 그림으로 제시하면, Table 4와 5 그리고 Figure 1과 같다.

잠재집단 a는 잠재집단 c보다 학교적응의 4가지 하위변수에서 유의하게 낮고, 잠재집단 b보다 학교적응의 3가지 하위변수, 즉 학교생활적응, 학업수행적응, 또래적응에서 유의하게 낮다. 따라서 잠재집단 a는 학교적응에 있어서 가장 낮은 수준을 보여 ‘낮은 수준의 학교적응’집단으로 명명하였다. 이 집단은 전체 아동의 13.1%를 차지하여 다른 두 집단에 비해 가장 적은 분포를 보인다.

잠재집단 b는 학교적응의 4가지 하위변수 모두가 잠재집단 c보다는 유의하게 작고 3가지 하위변수가 잠재집단 a보다는 유의하게 높다. 잠재집단 b는 학교적응에 있어서 중간 수준을 보여 ‘중간 수준의 학교적응’집단으로 명명하였다. 이 집단은

전체 아동의 28.7%를 차지하여 두 번째로 적은 분포를 보인다.

잠재집단 c는 잠재집단 a와 b보다 학교적응의 4가지 하위변수에서 유의하게 높다. 따라서 잠재집단 c는 학교적응에 있어서 가장 높은 수준을 보여 ‘높은 수준의 학교적응’집단으로 명명하였다. 이 집단은 전체 아동의 58.2%를 차지하여 가장 많이 분포한다.

Figure 1에서, 세 집단의 차이는 학교생활 적응에서 가장 크게 나타나며, 교사적응에 있어서는 가장 작은 차이를 보인다.

학교적응 잠재집단별 예측변수의 차이

잠재집단 유형별로 학교적응 예측변수에 유의한 차이가 있는지 분석한 결과는 Tabel 5와 같다.

Table 5

Comparison of Means of Predictors by Latent Groups

Sub variable	Class	<i>n</i>	<i>M</i>	<i>SD</i>	<i>F</i>	Scheffé
School readiness	a	132	3.45	0.43	10.72***	a = b < c
	b	290	3.51	0.37		
	c	588	3.59	0.32		
Difficulty in executive function	a	132	1.96	0.58	559.59***	a < b < c
	b	290	1.44	0.29		
	c	588	1.10	0.12		
School preference	a	132	3.02	0.94	7.12**	a < c a = b, c = b
	b	290	3.18	0.97		
	c	588	3.32	0.80		
Parent's interest in school life	a	132	3.73	1.00	37.40***	a = b < c
	b	290	3.92	0.86		
	c	588	4.30	0.74		
Teacher's teaching efficacy	a	132	4.30	0.52	32.17***	b < a < c
	b	290	4.17	0.45		
	c	588	4.43	0.44		
Teacher's work stress	a	132	3.17	0.86	7.51**	a = b < c
	b	290	3.18	0.77		
	c	588	2.97	0.86		

Note. *N* = 1,010. a = low level; b = middle level; c = high level.

p* < .01. *p* < .001.

Table 5에서, 잠재집단별로 유의한 차이가 있는 예측변수는 학교준비도, 집행기능 곤란, 학교선호도, 부모의 학교생활 관심도, 교사의 효능감과 업무스트레스이다. 사실상 대부분의 예측변수에서 집단별로 유의한 차이가 나타났다. 아동의 유아기 학교준비도는 저수준과 중수준 집단보다 고수준 집단에서 유의하게 높다($F = 10.72, p < .001$). 아동의 집행기능 곤란은 잠재집단의 3 집단이 모두 유의한 차이가 있었는데, 학교적응 저수준 집단의 집행기능 곤란이 가장 크고 학교적응 고수준 집단의 집행기능 곤란이 가장 작다($F = 559.59, p < .001$). 아동의 학교선호도는 학교적응 저수준 집단보다 고수준 집단에서 유의하게 높다($F = 7.12, p < .01$). 부모의 자녀 학교생활 관심도는 학교적응 저수준 및 중수준 집단보다 고수준 집단에서 유의하게 높다($F = 37.40, p < .001$). 교사의 효능감은 학교적응의 3 잠재집단 간에 유의한 차이가 있는데($F = 37.17, p < .001$), 특징적으로 중수준 집단보다 저수준 집단에서 교사의 효능감이 더 높다. 마지막으로 교사의 업무스트레스는 학교적응의 저수준 및 중수준 집단보다 고수준 집단에서 유의하게 낮다($F = 7.51, p < .01$).

머신러닝 모델별 예측력

아동의 학교적응 잠재집단 분류를 예측하기 위하여 7개의 예측변수를 4개의 머신러닝 모델에 투입하고 성능을 비교하였다(Table 6).

Table 6에서, 낮은 수준의 학교적응 집단을 나머지 집단과 구분하는 민감도는 로지스틱 회귀 모델, 라쏘 로지스틱 회귀 모델, 랜덤 포레스트 모델, 그레디언트-부스티드 의사결정 나무 모델 순서로 각각 .67, .64, .64, .70로 그레디언트-부스티드 의사결정 나무 모델이 가장 높다. 특이도는 4가지 모델 모두 .97로 매우 높다. 일반적으로 출현율이 작을수록 민감도가 낮아진다는 것을 고려할 때, 본 연구의 4가지 머신러닝 모델의 정확도는 .94로 높은 수준이다. 민감도와 특이도를 모두 고려한 AUC 값은 그레디언트-부스티드 의사결정 나무 모델이 .89로 가장 양호하며 로지스틱 모델보다 유의하게 크다($p < .05$). 그러나 나머지 3가지 분별모델(Low vs High, Low vs Middle, Middle vs High)에서는 랜덤 포레스트 모델이 전반적으로 성능이 가장 뛰어나다. 학교적응이 낮은 집단을 높은 집단과 분

Table 6

Machine Learning Model Performance Comparison Evaluation

	Model	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	AUC (95% CI)	<i>p</i>
Low vs Middle & High	Logistic	.94 (.91~.96)	.67 (.48~.82)	.97 (.95~.99)	.81 (.70~.93)	Reference
	Logistic with lasso	.94 (.90~.96)	.64 (.45~.80)	.97 (.95~.99)	.81 (.69~.93)	.291
	Random forest	.94 (.90~.96)	.64 (.45~.80)	.97 (.95~.99)	.88 (.81~.96)	.021
	Gradient-boosted	.94 (.91~.97)	.70 (.51~.84)	.97 (.94~.99)	.89 (.81~.96)	.010
Low vs High	Logistic	.96 (.91~.98)	.76 (.60~.88)	1.00 (.98~1.00)	.84 (.74~.93)	Reference
	Logistic with lasso	.95 (.91~.97)	.73 (.57~.86)	1.00 (.98~1.00)	.82 (.71~.92)	.019
	Random forest	.94 (.90~.97)	.76 (.60~.88)	.99 (.96~1.00)	.90 (.83~.97)	.078
	Gradient-boosted	.95 (.91~.97)	.73 (.57~.86)	1.00 (.88~1.00)	.83 (.73~.93)	.928
Low vs Middle	Logistic	.82 (.74~.88)	.55 (.38~.71)	.94 (.87~.98)	.71 (.59~.84)	Reference
	Logistic with lasso	.80 (.72~.87)	.53 (.36~.69)	.93 (.85~.97)	.70 (.57~.83)	.073
	Random forest	.80 (.71~.86)	.50 (.33~.67)	.93 (.85~.97)	.87 (.81~.94)	.010
	Gradient-boosted	.81 (.73~.88)	.61 (.43~.76)	.91 (.82~.96)	.84 (.75~.92)	.045
Middle vs High	Logistic	.81 (.76~.86)	.58 (.47~.69)	.92 (.87~.96)	.89 (.85~.93)	Reference
	Logistic with lasso	.82 (.76~.85)	.58 (.47~.69)	.93 (.88~.96)	.89 (.85~.93)	.566
	Random forest	.81 (.76~.86)	.60 (.48~.70)	.91 (.86~.95)	.86 (.81~.91)	.030
	Gradient-boosted	.80 (.75~.85)	.60 (.48~.70)	.91 (.86~.95)	.84 (.78~.89)	.035

Note. AUC = area under the ROC curve; *p* = significance level of Delong's test.

Figure 2. The ROC curve and decision curve of “Low versus Middle and High” model.

별할 때(AUC = .90, $p < .1$), 낮은 집단을 중간 집단과 분별할 때(AUC = .87, $p < .05$). 그리고 중간 집단을 높은 집단과 분별할 때(AUC = .89, $p < .05$), 랜덤 포레스트 모델의 AUC는 로지스틱 모델보다 유의하게 크며 대부분 우수한 수준에 가깝다.

Figure 2는 전체 아동에서 낮은 학교적응 집단을 분별할 때 (Low vs Middle & High), 네 가지 머신러닝 모델의 예측 능력을 시각적으로 보여준다. 먼저, 왼쪽 그림은 ROC curve로, 참

조모델인 로지스틱 회귀 모델에 비해 그레디언트-부스티드 의사결정 나무 모델의 면적이 더 크다는 것을 확인할 수 있다. 오른쪽 그림은 decision curve로 모델을 선택하는 이점을 나타낸다. 본 연구에서 낮은 학교적응 집단의 출현율(prevalence)은 세로선과 가로선이 만나는 13% 부근을 가리킨다. 4가지 머신러닝 모델의 순편익(net benefit)은 낮은 학교적응 집단에 모든 아동이 속한다는 가정(세로선)과 모든 아동이 속하지 않는다

Table 7

Predictor Coefficients in Logistic Regression Model

Variables	Model							
	Low vs Middle & High		Low vs High		Low vs Middle		Middle vs High	
	coef	OR	coef	OR	coef	OR	coef	OR
Gender (girls)	-0.02	0.07	-0.30	0.74	-0.17	0.84	-0.37**	0.69
School readiness	-0.02	0.98	0.16	1.17	0.05	1.05	0.00	1.00
Difficulty in executive function	1.65***	5.22	2.82***	16.86	1.37***	3.94	2.19***	8.91
School preference	-0.13	0.88	-0.06	0.95	-0.07	0.93	-0.17	0.85
Parent's interest in school life	-0.02	0.98	-0.15	0.87	0.05	1.05	-0.43***	0.65
Teacher's efficacy	0.32*	1.38	0.07	1.07	0.17	1.19	-0.54***	0.58
Teacher's work stress	0.32*	1.38	0.02	1.02	0.06	1.07	-0.05	0.96

Note. coef = regression coefficient of the variable; OR = odds ratio.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 3. The importance of variables of random forest model.

는 가정(가로선)보다 모두 높으며, 이러한 이점은 .80의 임계 값까지 유지된다.

학교적응 예측변수의 상대적 중요도

참조모델인 로지스틱 회귀모델에 총 7개의 변수를 투입하였을 때, 유의한 예측변수는 분별모델 별로 1-4가지 나타났다. 로지스틱 회귀 모델에서 예측변수가 아동의 학교적응 잠재집단 분류를 예측하는 정도는 회귀계수의 크기로 확인할 수 있

다(Table 7).

Table 7은 4가지 분별모델 각각의 로지스틱 회귀 모델에서 예측변수의 계수를 나타낸다. 투입변수의 단위는 모두 표준화되었기 때문에 회귀 모델의 계수크기는 표준화회귀계수처럼 상대적인 크기로 해석될 수 있다. 먼저, 전체 아동 중에서 낮은 학교적응 집단을 분별할 때, 낮은 학교적응 집단에 포함될 확률은 집행기능 곤란이 클수록 5.22배(coef = 1.65, $p < .001$), 교사의 효능감이 클수록 1.38배(coef = 0.32, $p < .05$), 교사의 업무스트레스가 클수록 1.38배(coef = 0.32, $p < .05$) 커진다. 교사의

효능감이 클수록 학교적응 저수준 집단에 포함될 확률이 커지는 것은 이 집단의 아동을 담당하는 교사의 효능감이 중수준 집단보다 높은 것과 관련 있다(Table 5). 다음으로, 낮은 학교적응 집단을 높은 학교적응 집단과 분별할 때, 낮은 학교적응 집단에 포함될 확률은 집행기능 곤란이 클수록 16.86배 높다(coef = 2.82, $p < .001$). 낮은 학교적응 집단을 중간 수준의 학교적응 집단과 분별할 때에도, 낮은 학교적응 집단에 포함될 확률은 집행기능 곤란이 클수록 3.94배 높다(coef = 1.37, $p < .001$). 마지막으로, 학교적응 중간 수준의 집단을 높은 학교적응 집단과 분별할 때, 학교적응 중간 집단에 포함될 확률은 아동의 집행기능 곤란이 클수록 8.91배 증가하고(coef = 2.19, $p < .001$), 여아일수록 0.69배로 감소하며(coef = -0.37, $p < .01$), 부모의 학교생활 관심도가 클수록 0.65배로 감소하고(coef = -0.43, $p < .001$), 교사의 효능감이 클수록 0.58배(coef = -0.54, $p < .01$)로 감소한다.

Figure 2는 4가지 학교적응 분별모델에 대해서 전반적으로 양호한 성능을 보였던 랜덤포레스트 모델에 예측변수의 중요도를 상대적으로 시각화한 것이다. 가장 기여도가 높은 변수의 중요도를 100으로 가정하고 비교할 때, 4가지 분별모델에서 아동의 집행기능 곤란이 모두 100으로 학교적응에 가장 중요한 변수이다.

Discussion

본 연구는 한국아동패널 7·8차 자료를 활용하여 우리나라 초등학교 1학년의 학교적응 유형을 잠재프로파일분석으로 분류하고, 머신러닝 기법 중 로지스틱 회귀, 라쏘 로지스틱 회귀, 랜덤 포레스트, 그레디언트-부스티드 의사결정나무 모델을 적용하여 아동의 학교적응 유형을 예측하고 분별하는 모델을 개발하였다. 연구결과를 토대로 다음과 같이 논의하였다.

첫째, 학교적응 유형을 분류하는 잠재프로파일분석에서, 학교적응의 잠재집단은 학교적응 수준이 ‘높은 수준’(58.2%), ‘중간 수준’(28.7%), ‘낮은 수준’(13.1%)으로 명명된 3개의 유형이 나타났다. 한국아동·청소년패널 자료를 활용한 Park (2019)의 연구에서 초등학교 6학년의 학교적응에 대한 잠재프로파일은 가장 높은 적응력을 보이는 학교친화형(10.1%), 적응형(57.3%), 학교적응이 가장 낮은 부적응형(32.6%)의 3가지 유형으로 나타났는데, 본 연구에서와 같이 학교적응에 있어 부적응형에 비해 적응형의 비율이 더 높다. 그러나 고학년 아동의 자료를 분석한 Park (2019)의 연구보다 1학년 아동의

자료를 분석한 본 연구에서 학교적응이 양호한 아동의 비율이 더 높다. 초등학교 저학년과 고학년의 학교적응 요인을 비교한 연구에 의하면, 아동의 학업성취 수준과 부모의 방임은 저학년과 고학년 모두에게 학교적응에 영향을 미치는 공통적인 예측요인이었다(K. Lee & Lee, 2013). 또한, 저학년 아동에게는 특징적으로 건강상태가 학교적응의 예측요인이었는데, 이러한 결과는 초등학교 저학년 아동이 스트레스에 대한 반응으로 신체화 증상이 나타나지만, 초등학교 고학년과 중·고등학생은 반항, 비행과 같은 외현적인 행동으로 나타난다(Lim & Woo, 2017)는 사실과 관련지어 볼 수 있다. 아동의 신체건강 상태가 스트레스와 같은 정신건강 상태와 관련이 있는지 추후 확인할 필요가 있지만, 저학년 시기에도 학업성취에 대한 압력은 존재하며 고학년이 될수록 이러한 압력은 점점 더 커져(Hong & Lee, 2012) 이에 대해 거부와 반발은 아동의 행동문제로 표출될 수 있다(J. Lee et al., 2015). 이러한 사실들을 고려하면, 초등학교 1학년 시기의 학교적응 수준은 2, 3학년에 비해 낮지만, 고학년과 비교해서는 높은 수준이 될 것이다. 이러한 1학년 시기의 학교적응 특징은 초등학교 입학 직후라는 특수한 상황을 반영한다.

둘째, 4가지 머신러닝 모델(로지스틱 회귀 모델, 라쏘 로지스틱 회귀 모델, 랜덤 포레스트, 그레디언트-부스티드 의사결정나무 모델)은 AUC 값을 기준으로 할 때 모두 .80 이상으로 양호하거나, .90에 가까운 우수한 예측력을 보였다. 특히, 본 연구의 머신러닝 모델 중 랜덤 포레스트 모델의 예측력은 참조모델인 로지스틱 회귀 모델보다 유의하게 예측력이 높았다. 즉, 하이퍼-파라미터의 제어가 없는 기본적인 회귀모델에 비해 랜덤 포레스트 모델과 같은 머신러닝 모델이 아동의 학교적응을 더 정확하게 분류하였다. 따라서 한국아동패널자료를 활용한 머신러닝 기법이 개별 아동의 학교적응 수준을 분류하는 데 유용함을 본 연구결과로 확인하였다.

셋째, 4가지 머신러닝 기법별로 초등학교 1학년 아동의 학교적응의 유형을 예측하는 중요도의 양상은 대체로 유사하였다. 머신러닝 모델 중에서 랜덤 포레스트 모델이 4가지 분별모델(Low vs Middle & High, Low vs High, Low vs Middle, Middle vs High)에 대한 성능이 가장 양호하여, 이 모델로 예측변수의 상대적 중요도를 비교하였다. 분석 결과, 학교적응 유형을 예측하는 가장 중요한 변수는 4가지 분별모델에서 공통적으로 ‘집행기능 곤란’이었으며, 집행기능 곤란의 상대적인 중요도도 다른 변수의 중요도에 비해 압도적으로 높았다. 본 연구에서 집행기능 곤란의 하위영역은 주의, 정서통제, 행동통제, 계획·조직화에서 곤란이었다. 1학년 아동의 학교적응 프로파일

은 착석과 같은 질서 및 규칙과 관련된 학교생활 적응에서 잠재집단간 격차가 가장 컸다. 이러한 질서와 규칙을 준수하기 위해서는 아동 자신의 주의와 행동을 통제하는 능력이 필요하다는 점에서 집행기능 곤란은 학교생활 적응에 강력한 위험요인이 될 수 있다. 또한, 학교생활 하위영역 중에서 학업성취와 또래 관계도 세 집단 간 격차가 유의하게 나타났는데, 초등학교 저학년 시기 동안 학업성취를 지속해서 강화하는 요인은 집행기능이었다(Morgan et al., 2019). 그리고 정서조절이 또래 관계에 정적인 영향을 미친다(Jang & Lee, 2011)는 점에서도 집행기능의 정서통제 곤란은 아동의 학교적응을 어렵게 한다. 그동안 일부 연구에서 학교적응에 대한 집행기능의 영향을 보고하였지만(Yi & Jun, 2019), 한국아동패널자료를 이용하여 초등학교 1학년의 학교적응을 다면적으로 예측한 선행연구에서는 집행기능의 중요성은 부각되지 못했다(C.-I. Kim, 2018; E. Kim, 2018; J. Kim & Hong, 2019a; W. Kim & Kwon, 2019). 그러나 본 연구결과에서와 같이 집행기능 곤란의 부정적 영향을 고려할 때 이를 지원할 방안을 모색할 필요가 있다.

집행기능 곤란 이외에 아동의 학교적응을 예측하는 변수로는 먼저, 아동의 성별에서 여학생일수록 전체 아동집단에서 낮은 수준의 학교적응 집단을 분별하는데 기여하였다. 이는 여아가 남아보다 초기 학교적응에 유리하다는 결과(C.-I. Kim, 2018; E. Kim, 2018; J. Kim & Hong, 2019a; Roeser & Eccles, 1998; Simons-Morton & Crump, 2003)를 지지하는 결과이다. 따라서 여아보다 남아의 학교적응을 위한 관심과 지원이 더 필요할 수 있다. 다음으로, 교사의 효능감과 업무스트레스도 전체 아동집단에서 낮은 수준의 학교적응 집단을 분별하는데 기여하였다. 특징적으로, 낮은 수준의 학교적응 집단을 담당하는 교사의 효능감이 중간 수준의 학교적응 집단을 담당하는 교사보다 유의하게 높았다. 본 연구결과는 교사의 효능감이 높을수록, 업무스트레스는 낮을수록 아동의 긍정적인 학교적응을 예측한다는 선행연구(J. Kim & Hong, 2019a)를 기본적으로 지지하지만 이러한 결과가 나타난 것은 교수 경험의 결과로 이해할 필요가 있다. 본 연구의 교사효능감 측정 자료는 교사가 아동의 학업뿐만 아니라 일상생활과 행동문제까지 지도하는 내용을 포함하는데, 학교적응에 어려움이 있는 아동들을 효과적으로 지도하면서 교사의 효능감은 오히려 향상될 수 있기 때문이다. 따라서 학교적응에 어려움을 겪는 아동을 담당하는 교사를 지원하는 것은 교사뿐만 아니라 교사가 이후에 담당할 아동을 지원하는 방안이 될 수 있다. 한편, 본 연구에서 부모의 자녀 학교생활 관심도는 높은 학교적응 집단을 다른 집단과 구별하는데 기여하는 것으로 나타나, 부모의 사랑과

관심이 아동의 학교적응에 긍정적으로 영향을 미친다는 선행연구(E. Kim, 2018; J.-H. Lee & Kim, 2020)를 지지한다. 따라서 부모가 초등학교에 입학한 자녀의 학교생활에 관심을 가질 수 있도록 부모의 참여를 유도하는 것이 아동의 학교적응을 돕는 지원책이 될 수 있다.

그러나 아동의 성별, 교사의 효능감과 부모의 관심은 아동의 긍정적인 학교적응을 예측하는 유의한 변수였지만, 이들 변수의 중요도를 집행기능 곤란과 비교하면 압도적으로 작았다. 특히, 유아기 학교준비도가 초등학교 입학 이후 학교적응에 영향을 미친다는 선행연구(Y. A. Jung & Kim, 2019; J. Kim, 2019; J. Kim & Hong, 2019b; Y. H. Kim, Kim, & Kim, 2019; S. A. Lee & Lim, 2019)와는 달리 본 연구에서 유아기 학교준비도는 아동의 학교적응 수준을 분류하는 데 중요하게 기여하지 못했다. 이는 초등교육으로의 전이를 유아교육에서는 ‘준비’로 바라보는 관점과 초등교육에서 ‘적응’으로 바라보는 관점 간의 격차와 관련있다(Jeong & Park, 2017). 학교준비도는 아동발달의 전인적 측면을 포괄하고 있으며, 이러한 전인적 발달을 성취하기 위한 노력은 아동중심적인 교육관과 문화적 풍토 속에서 이루어진다. 그러나 초등학교에서의 적응은 비록 학교생활, 학업수행, 인간관계의 다양한 측면을 다루고 있지만 각 영역에서 아동의 자기조절 능력을 강하게 요구한다. 특히, 본 연구에서 아동의 학교적응과 집행기능 곤란은 모두 교사가 측정하는 자료라는 점에서 교사가 인식하는 학교적응은 집행기능과 밀접한 관련이 있다. 그러나 아동의 자기조절에 대한 갑작스런 기대는 유아에게 문화충격으로 다가올 수 있다. 따라서 아동의 성공적인 초등학교 전이를 위해서는 유아교육과 초등교육의 간극을 극복할 수 있도록 아동을 준비시키는 것뿐만 아니라 교실의 문화적 간극을 줄이려는 교사, 학교, 정책 차원의 노력도 필요하다.

본 연구의 의의를 서술하면, 첫째, 본 연구는 1학년 아동의 학교적응에 대한 잠재프로파일분석을 실시하여 학교적응에서 상이한 양상을 보이는 집단을 규명하였으며, 머신러닝 기법을 활용하여 기존의 회귀분석 연구가 보여주었던 인과적 추정뿐 아니라 개별 아동을 분류하는 유용성도 보여주었다. 이를 통해 본 연구에서 개발된 모델은 아동의 학교적응을 돕는 보편적 지원뿐 아니라 아동 개개인의 선별과 맞춤형 지원을 위한 근거로 활용될 수 있을 것이다. 둘째, 본 연구는 초등학교 입학 후 학교적응에 있어 집행기능 곤란이 갖는 강한 영향력을 밝힌 의의가 있다. 기존의 연구에서는 초등학교 연계를 위해 학교준비도를 강조하였지만, 초등학교 입학 후 실제적인 학교적응을 위해서는 교실문화의 점진적인 변화와 아동의 집

행기능 증진을 지원하는 것이 중요함을 밝힌 의의가 있다. 셋째, 본 연구는 한국아동패널자료의 대규모 종단자료를 활용하여 이질적인 두 시기의 요인까지 고려하고 정확도가 높은 분별모델을 개발했다는 점에서 머신러닝 기법의 유용성뿐만 아니라 한국아동패널 자료의 유용성도 확인한 의의가 있다.

이러한 의의에도 본 연구는 몇 가지 제한점을 가지고 있어 후속 연구를 위한 제안을 하고자 한다. 첫째, 본 연구에서 분석에 활용한 자료는 초등학교 1학년의 자료만을 활용한 한계가 있다. 후속 연구에서는 잠재전이분석과 잠재계층성장분석 등을 활용하여 초등학교의 전 과정과 중학교로의 전이 과정에서 학교적응에 기여하는 예측요인들을 종단적으로 규명할 필요가 있다. 둘째, 본 연구에서는 학교적응을 집행기능 곤란에 예측하는 데 있어 두 변수의 총점을 투입하여 분석하였다. 그러나 이러한 분석은 학교적응의 각 하위요인에 집행기능 곤란의 하위요인들이 어떻게 기여하는지 확인할 수 없다. 또한, 본 연구의 집행기능 곤란은 교사가 측정한 자료로서 부모가 측정된 자료로 분석한 결과와 비교할 필요가 있다. 셋째, 본 연구의 머신러닝 모델에서는 각각의 변수들이 갖는 중요성의 크기를 비교할 뿐, 각 변수의 구조적인 관련성을 보여주지 못한다. 후속 연구에서는 다양한 구조분석을 통해 이들 변수의 역동적인 관계를 확인할 필요가 있다.

Acknowledgements

This study was supported by the 2020 research fund of Dong-A University.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

References

In English

- Barkley, R. A., & Lombroso, P. J. (2000). Genetics of childhood disorders: XVII. ADHD, Part 1: The executive functions and ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(8), 1064-1068. doi:10.1097/00004583-200008000-00025
- Bergman, L. R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9(2), 291-319. doi:10.1017/s095457949700206x
- Breiman, L. (2001) Random forests. *Machine Learning*, 45, 5-32. doi:10.1023/A:1010933404324
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3), 837-845. doi:10.2307/2531595
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802-813. doi:10.1111/j.1365-2656.2008.01390.x
- Hill, N. E., & Tyson, D. F. (2009). Parental involvement in middle school: A meta-analytic assessment of the strategies that promote achievement. *Developmental Psychology*, 45(3), 740-763. doi:10.1037/a0015362
- Jervis, K., & Polland, B. K. (2007). *Separation: Supporting children in their preschool transitions*. Retrieved from ERIC database. (ED498860)
- Kagan, S. L., & Kauerz, K. (2007). Reaching for the whole: Integration and alignment in early education policy. In R. C. Pianta, M. J. Cox, & K. L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability* (pp. 11-30). Baltimore, MD: Paul H Brookes Publishing.
- Korkman, M., Kemp, S. L., & Kirk, U. (2001). Effects of age on neurocognitive measures of children ages 5 to 12: A cross-sectional study on 800 children from the United States. *Developmental Neuropsychology*, 20(1), 331-354 doi:10.1207/S15326942DN2001_2
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. New York, NY: Springer.
- Masten, A. S., Herbers, J. E., Desjardins, C. D., Cutuli, J. J., McCormick, C. M., Sapienza, J. K.,...Zelazo, P. D. (2012). Executive function skills and school success in young children experiencing homelessness. *Educational Researcher*, 41(9), 375-384. doi:10.3102/0013189x12459883
- Morgan, P. L., Farkas, G., Hillemeier, M. M., Pun, W. H., & Maczuga, S. (2019). Kindergarten children's executive functions predict their second-grade academic achievement and behavior. *Child Development*, 90(5), 1802-1816. doi:10.1111/cdev.13095
- Muthukrishnan, R., & Rohini, R. (2016, October). LASSO: A feature selection technique in predictive modeling for machine learning. *IEEE International Conference on*

- Advances in Computer Applications (ICACA)*, Coimbatore, India. doi:10.1109/ICACA.2016.7887916
- Quirk, M., Grimm, R., Furlong, M. J., Nylund-Gibson, K., & Swami, S. (2016). The association of Latino children's kindergarten school readiness profiles with grade 2-5 literacy achievement trajectories. *Journal of Educational Psychology*, 108(6), 814-829. doi:10.1037/edu0000087
- Roeser, R. W., & Eccles, J. S. (1998). Adolescents' perceptions of middle school: Relation to longitudinal changes in academic and psychological adjustment. *Journal of Research on Adolescence*, 88(1), 123-158.
- Safari, S., Baratloo, A., Elfil, M., & Negida, A. (2016). Evidence based emergency medicine; Part 5 receiver operating curve and area under the curve. *Emergency (Tehran, Iran)*, 4(2), 111-113.
- Simons-Morton, B. G., & Crump, A. D. (2003). Association of parental involvement and social competence with school adjustment and engagement among sixth graders. *Journal of School Health*, 73(3), 121-126. doi:10.1111/j.1746-1561.2003.tb03586.x
- Vickers, A. J., & Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Medical Decision Making*, 26(6), 565-574. doi:10.1177/0272989X06295361
- Weiten, W., Lloyd, M. A., Dunn, D., & Hammer, E. Y. (2009). *Psychology applied to modern life: Adjustment in the 21st century* (9th ed.). Belmont, CA: Cengage Learning.
- Wyrick, A. J., & Rduasill, K. M. (2009). Parent involvement as a predictor of teacher-child relationship quality on third grade. *Early Education and Development*, 20(5) 845-864. doi:10.1080/10409280802582803
- Jang, M. H., & Lee, J. Y. (2011). Effects of mother attachment, rejection sensitivity, and emotional regulation on children's peer relationships. *Korean Journal of Play Therapy*, 14(1), 35-53.
- Jeong, K., & Park, C. (2017). Exploring viewpoints of transition to elementary school. *Journal of Curriculum Integration*, 11(4), 121-143.
- Jung, Y. A., & Kim, S. J. (2019). School readiness profiles on school adjustment in elementary school. *Journal of Learner-Centered Curriculum and Instruction*, 19(18), 875-895. doi:10.22251/jlcci.2019.19.18.875
- Jung, M. Y., & Moon, H. J. (2007). Relevant variables of children's school adjustment. *Korean Journal of Child Studies*, 28(5), 37-54.
- Kim, C.-I. (2018). The relationship between socio demographic factors and school adjustment for a first-year elementary school student. *The Journal of Child Education*, 27(4), 157-172. doi:10.17643/KJCE.2018.27.4.08
- Kim, E. (2018). Exploring effective variables for school adjustment in the first grade. *The Korean Journal of Applied Developmental Psychology*, 7(1), 19-34. doi:10.22839/adp.2018.7.1.19
- Kim, J. (2019). An analysis of exploring the effects of school readiness on school adjustment and academic achievement. *Journal of Learner-Centered Curriculum and Instruction*, 19(8), 1111-1130. doi:10.22251/jlcci.2019.19.8.1111
- Kim, J., & Hong, S. (2019a). An analysis of factors affecting the school adjustment in elementary school students. *Journal of Learner-Centered Curriculum and Instruction*, 19(6), 699-719. doi:10.22251/jlcci.2019.19.6.699
- Kim, J., & Hong, S. (2019b). Examining structural relationships among school readiness, academic achievement, self-esteem, and school adjustment. *Journal of Education & Culture*, 25(3), 507-525. doi:10.24159/joec.2019.25.3.507
- Kim, J. Y., & Lee, K. (2020). The effects of mother's educational involvement on first-grader's school adjustment: The interactive effects of teacher-parents relationship. *Journal of Parent Education*, 12(1), 81-103. doi:10.36431/JPE.12.1.5
- Kim, M. (2008). An exploratory study of student's experience of transition from kindergarten to first grade. *Gjyoyukgwahagyeongu [교육과학연구]*, 39(1), 181-210. <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01781598>
- Kim, S. (2018). Trajectories of externalizing behavior problems in young children and its associations with elementary school adjustment and executive function difficulties: Using growth mixture modeling. *Korean Journal of Child Studies*, 39(6), 41-56. doi:10.5723/kjcs.2018.39.6.41
- Kim, Y. H., Kim, H. K., & Kim, J. H. (2019). Longitudinal effects of children's temperamental profiles on school readiness and school adjustment: A person-centered approach. *Korean Journal of Early Childhood Education*, 39(6), 377-400.

In Korean

- Chi, S.-A., & Jung, D.-H. (2006). Validation of school adjustment inventory for first grade elementary school students. *Korean Journal of Child Studies*, 27(1), 1-15.
- Hong, Y. J., & Lee, J. Y. (2012). The mediating effects of parental learning involvement and academic self-efficacy of elementary school students between parental academic achievement pressure and academic achievement. *The Korean Journal Child Education*, 21(2), 325-342.
- Hwang, Y. S. (2012). A research on the articulation between early childhood curriculum and elementary school. curriculum. *Korean Journal of Early Childhood Education*, 32(5), 321-345. doi:10.18023/kjece.2012.32.5.015
- Im, H., Kim, Y., Shin, H., Lee, K., Jang, Y., & Kee, K. (2010). *Analysis on the actual status and qualities of school education in Korea: A study on the middle schools (III)* (Report No. RR 2010-23). JinCheon, Chung-Buk: Korean Educational Development Institute.

- doi:10.18023/kjece.2019.39.6.016
- Kim, W., & Kwon, H. (2019). Effects of psychological, socio-environmental, and mothers' parenting behavior factors on school adjustment in Korea children. *The Korean Journal of Developmental Psychology*, 32(4), 1-19. doi:10.35574/KJDP.2019.12.32.4.1
- Kwon, Y. H., Jeon, H.-J., & Goh, E. K. (2019, November). Adonghakoeji 40nyeonui yeoksa: Baldal [아동학회지 40년의 역사: 발달]. Paper presented at the 2019 Annual Autumn Conference of the Korean Association of Child Studies, Seoul, Korea.
- Lee, J., Kang, S., Lee, K., & Yi, S. (2015). The effects of parenting practice on the trajectories of change in the school adjustment of upper elementary students. *Korean Journal of Child Studies*, 36(4), 91-110. doi:10.5723/KJCS.2015.36.4.91
- Lee, J.-H., & Kim, D.-W. (2020). Effects of mother's employment and work-rearing conflict on elementary school adaptation after kindergarten. *Early Childhood Education Research & Review*, 24(3), 263-283. doi:10.32349/ECERR.2020.6.24.3.263
- Lee, K. (2017). The effects of gender, resilience, study habit, and parenting practice on the trajectories of change in the school adjustment of upper elementary students. *Journal of Learner-Centered Curriculum and Instruction*, 17(1), 295-322. doi:10.22251/jlcci.2017.17.1.295
- Lee, K. (2020). Causal longitudinal analysis between media usage time trajectories and school adaptation trajectories for low grade elementary school children. *Journal of Parent Education*, 12(1), 105-124. doi:10.36431/JPE.12.1.6
- Lee, K., & Lee, M. (2013). Effects of personal, family, and school variables on school adjustment among elementary school and middle school students. *Journal of Life-span Studies*, 3(1), 13-28.
- Lee, S. A., & Lim, S. A. (2019). The mediated effects of school readiness and social skills in the relationship between children's developmental characteristics and first graders' school adjustment. *Korean journal of elementary education*, 30(1), 221-233. doi:10.20972/Kjee.30.1.201903.221
- Lim, W.-S., & Woo, G. I. (2017, November 15). Chodeungsaengttaen boktong deung hoso... junggosaeng ollagameon banhang-bihaeng pokbal [초등생땀 북통 등 호소... 중고생 올라가면 반향-비행 폭발]. *The Dong-A Ilbo*. Retrieved from <https://www.donga.com/news/Society/article/all/20171115/87265104/1>
- No, B., Park, S., Park, H. J., & Choi, N. (2020). An analysis of latent profiles of educational support in parents of early school-aged children: Predictors and differences in children's school adjustment. *Korean Journal of Child Studies*, 41(1), 15-28. doi:10.5723/kjcs.2020.41.1.15
- Panel Study on Korean Children. (2014). *Panel study of Korean children 7th survey* [Data file and codebook]. Retrieved from <https://panel.kicce.re.kr/panel/module/rawDataManage/index.do>
- Panel Study on Korean Children. (2015). *Panel study of Korean children 8th survey* [Data file and codebook]. Retrieved from <https://panel.kicce.re.kr/panel/module/rawDataManage/index.do>
- Park, S. (2019). Latent change of school adjustment profiles and its predictors across the transition from primary to middle school. *Journal of Learner-Centered Curriculum and Instruction*, 19(21), 821-841. doi:10.22251/jlcci.2019.19.21.821
- Song, H. (2014). Validity of child-adolescent self-reported executive function difficulty screening questionnaire. *The Korean Journal of Clinical Psychology*, 33(1), 121-137. doi:10.15842/kjcp.2014.33.1.008
- Sung, K.-S., Kim, J., Park, S.-Y., & Min, B.-C. (2012). *A study of research design and item development for Gyeonggi education panel study*. Suwon, Korea: Gyeonggi Institute of Education.
- Sung, M., Chang, Y. E., & Seo, B. (2016). The roles of study habits and emotional-behavioral problems in predicting school adjustment classification among 3rd graders. *Korean Journal of Childcare and Education*, 12(6), 79-102. doi:10.14698/jkcce.2016.12.06.079
- Yi, Y., & Jun, E. (2019). Exploring variables affecting school adjustment in first grade elementary students: Focused on mother's employment, home environment, and children's executive function difficulty. *Korean Journal of Child Care and Education Policy*, 13(1), 101-121. doi:10.5718/kcep.2019.13.1.101

ORCID

Hyo Jeong Jeon <http://orcid.org/0000-0002-9714-7599>
 Eun Kyoung Goh <http://orcid.org/0000-0003-4811-3664>

Received June 30, 2020

Revision received July 27, 2020

Accepted August 11, 2020