

머신러닝을 활용한 청소년 독서 예측: 주요 영향 요인 분석과 모델 비교*

Predicting Youth Reading Behavior Using Machine Learning: Analysis of Key Factors and Model Comparison

신인수 (Insoo Shin)**

박성재 (Sungjae Park)***

초 록

본 연구의 목적은 머신러닝 기법을 적용하여 중학생들의 독서 여부에 영향을 미치는 요인을 분석하고 최적의 기법을 찾는 것이다. 2018 한국아동·청소년패널 데이터의 독서시간 변수를 이용하여 주중 독서시간이 1시간 이상인 학생을 독서집단으로, 독서를 하지 않은 학생을 비독서집단으로 구분하고 머신러닝 기법으로 로지스틱 회귀, 의사결정나무, 랜덤 포레스트, adaptive LASSO, SVM, 그라디언트 부스팅, kNN을 적용하여 비교 분석하였다. 분석 결과, 랜덤 포레스트와 그라디언트 부스팅이 가장 높은 예측 정확도를 보였다. 앙상블 기법은 비선형적 패턴을 효과적으로 포착하여 독서 여부를 정교하게 분류함을 보여주었다. 부분의존성 도표를 통해 학업 열의, 친구 및 교사 관계, 스마트폰 의존, 자아존중감 등이 독서 결정에 중요한 변수임이 확인되었다. 학업 열의와 자아존중감이 높을수록 독서 확률이 상승하는 한편, 학업무기력과 스마트폰 의존은 독서를 저해하였다. 본 연구 결과는 독서 장려 정책에 필요한 요인을 구체적으로 제시한다는 점에서 의의가 있다. 특히, 학업 동기부여, 친구 및 교사 지원, 미디어 활용 교육 등을 연계한 다각적 접근의 필요성을 시사한다.

ABSTRACT

The purpose of this study is to identify the factors influencing middle school students' reading behavior by applying various machine learning methods and to determine which method yields the best performance. Using data from the 2018 Korean Child and Youth Panel Survey, students who read more than one hour per weekday were classified as the reading group, while those who did not read at all were classified as the non-reading group. Seven machine learning methods-*logistic regression, decision tree, random forest, adaptive LASSO, SVM, gradient boosting, and kNN*-were then applied for comparative analysis. The analysis shows that random forest and gradient boosting achieved the highest predictive accuracy, indicating that ensemble methods can effectively capture nonlinear patterns and distinguish reading behavior more precisely. According to the partial dependence plots, academic engagement, relationships with peers and teachers, smartphone dependence, and self-esteem were significant variables in determining whether students read. While stronger academic engagement and higher self-esteem increased the likelihood of reading, academic helplessness, and excessive smartphone use impeded it. This study is significant in that it provides detailed factors necessary for designing effective reading promotion policies. In particular, it underscores the need for a multifaceted approach that integrates academic motivation strategies, peer and teacher support, and media utilization education.

키워드: 독서, 청소년, 머신러닝, 랜덤 포레스트, 그라디언트 부스팅, 한국아동·청소년패널조사
reading, youth, machine learning, random forest, gradient boosting, the Korean Child and Youth Panel Survey

* 본 연구는 한성대학교 학술연구비 지원과제임.

이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2021S1A3A2A01090926).

** 동국대학교 교육대학원 인공지능융합교육 교수(s9065031@dongguk.edu) (제1저자)

*** 한성대학교 지식정보문화트랙 교수(spark@hansung.ac.kr) (교신저자)

■ 논문접수일자: 2025년 2월 16일 ■ 최초심사일자: 2025년 3월 4일 ■ 게재확정일자: 2025년 3월 4일

■ 정보관리학회지, 42(1), 131-153, 2025. <http://dx.doi.org/10.3743/KOSIM.2025.42.1.131>

※ Copyright © 2025 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

독서의 중요성이 강조되는 요즘 역설적이게도 청소년들의 독서율은 지속적으로 하락하고 있다(문화체육관광부, 2023). 독서율을 제고하기 위해서는 여러 접근법이 있다. 먼저 독서를 하지 않는 이유를 파악하여 그 원인을 제거하거나 통제함으로써 독서로 이끄는 것이다(이경남, 2016; Gallagher, 2023). 또 다른 하나는 독서를 하는 청소년들은 어떤 환경에서 어떤 요인에 의해 영향을 받는지를 분석하여 유사한 환경을 만듦으로써 독서율을 끌어 올리는 것이다(김영식 외, 2019; 남수정, 2023; 박성재, 2022; 조중열, 이임숙, 2004). 본 연구는 후자의 접근법을 통해 독서율을 제고하기 위한 방안을 연구하고자 한다.

독서에 영향을 미치는 예측 변수를 탐구하는 연구는 많은 연구자들의 관심 주제 중의 하나이다. 독서에 대한 선행 연구나 이론을 바탕으로 하거나 다른 분야, 특히 인간행동에 대한 연구에서 사용하는 모델이나 이론을 적용하여 요인을 분석했다(변우열, 2012; 이병기, 2015). 이러한 연역적 접근법은 튼튼한 이론적 기반으로 독서와 관련 변인들 간의 관계를 증명할 수 있다는 점에서 의미가 있다. 그러나 기존의 연구에서 다루어졌던 독서 관련 변인들로 제한된다는 점에서 다양한 변인들을 고려하지 못한다는 한계점이 있다(박성재, 신인수, 2024). 이러한 한계점을 극복하기 위한 방법으로 탐색적·귀납적 연구가 최근 활발히 진행되고 있다. 특히, 데이터와 이를 처리하는 컴퓨터 성능 향상과 알고리즘의 개발은 귀납적 접근을 용이하게 하였다.

독서 연구 분야에서도 탐색적·귀납적 접근법을 적용한 연구가 진행되고 있다. 최근 인공지능 기술, 특히 머신러닝 및 딥러닝 기법이 발전함에 따라 과학기술 분야뿐만 아니라 인문·사회과학 분야에서도 최신 기술을 적용한 연구를 수행하고 있다. 머신러닝과 관련하여 연구 분야에서 관심을 갖는 대표적인 알고리즘에는 로지스틱 회귀, 의사결정나무(decision tree), 랜덤 포레스트(random forest), SVM, 그래디언트 부스팅(gradient boosting), SVM(Support Vector Machine), adaptive LASSO(adaptive Least Absolute Shrinkage and Selection Operator), kNN 등이 있다. 김영식 외(2019)는 랜덤 포레스트 기법을 적용하여 독서에 영향을 미치는 주요 요인을 분석하였다. 그러나 앞서 언급한 것과 같이 머신러닝 기법에는 여러 가지가 있고 각각은 주제와 데이터의 특성에 따라 다른 연구 결과를 보인다는 점에서 적합한 모형을 찾기 위한 연구가 선행될 필요가 있다.

따라서 본 연구는 주요 머신러닝 기법을 적용하여 청소년들의 독서에 영향을 미치는 요인이 무엇인지, 그리고 가장 예측력이 좋은 기법이 무엇인지를 분석하고자 한다. 이러한 연구 목적을 바탕으로 다음과 같은 연구 문제를 설정하였다.

- 연구문제 1. 머신러닝 기법 중 청소년의 독서 영향 요인을 가장 잘 예측하는 기법은 어느 것인가?
- 연구문제 2. 독서에 영향을 미치는 요인은 무엇인가?

2. 이론적 배경

2.1 독서에 영향을 미치는 요인

독서에 영향을 미치는 요인은 크게 심리적 요인, 사회적 요인, 환경적 요인으로 나눌 수 있다. 먼저 심리적 요인은 독서와 관련된 개인의 심리적인 상태를 나타내며 독서 태도, 흥미, 습관, 선호 등을 주제로 한 연구이다. 정혜승(2006)은 독서 태도를 “독서를 대상으로 좋고 싫음의 평가를 내리고 그러한 평가에 따라 독서 상황에 접근하거나 회피하는 등의 반응을 일관되게 산출하는 심리적 경향”으로 정의하였다. 왕효성(2014)은 독서 태도에 영향을 미치는 요인을 분석했는데 관심 및 흥미(50.8%), 인지적 보상(11.8%), 몰입 독서(10.25), 정서적 보상(8.3%) 등이 주요한 요인으로 나타났다. 특히, 몰입과 관련하여 칙센트미하이의 몰입이론(Flow Theory)을 적용하여 독서 몰입을 통해 독서 태도, 흥미, 습관이 형성됨을 보여주고 있다. 더 나아가 몰입을 통해 독서동기화는 물론 독자로서의 자기인식과 자기효능감 향상을 가져왔다는 연구결과도 있다(Myers, 1992; Schallert & Reed, 1997). 또한 내재적 독서동기가 높은 학생들은 독서량(Baker & Wigfield, 1999)과 독서 빈도(Wigfield & Guthrie, 1995)에서도 높게 나타났다. 심리적인 요인 중의 하나인 독서 선호와 관련하여 독서 동기가 높은 학생일수록 선호가 높게 나타났다(박혜숙, 2008). 선호를 심층적으로 분석한 연구에서는 선호가 낮은 사람들에게는 독서를 하는 방법을 안내함으로써, 선호가 높은 사람들은 독서의 이유와 목적을 안내함으로써 독서로 이끌 수 있다고 제시하고 있

다(이국희 외, 2020). 그러나 독서 선호는 심리적인 바램을 나타낼 수 있다는 점에서 행위적인 관점에서 독서 습관이 중요하고(이국희, 2022) 마음 먹은 것을 실행할 수 있는 실행기능 전환능력이 독서에서도 중요하다(김아란 외, 2021)

사회적 요인에 대한 연구는 독자 개인보다는 다른 구성원과의 관계가 개인의 독서에 영향을 미친다는 점을 강조하고 있다. 사회적 관계가 형성되는 가족이나 학교, 지역사회에서 타인과의 관계 맺음과정에서 개인의 독서가 확대되거나 축소될 수 있다는 것이다. 독서 시간에 영향을 미치는 가족과 관련된 변인으로 어머니의 학력(하문선, 2017), 가구소득(김영식 외, 2019), 사회경제적 지위(박소현, 2020) 등이 중요한 요인으로 나타났다. 학교에서의 친구들과의 관계 또한 독서에 영향을 미치고 있었으며 독서하는 친구를 두고 있는 학생들은 독서를 할 확률이 독서를 하지 않는 학생들보다 높았다(박성재, 신인수, 2024). 학교 환경에서 교사와의 관계 또한 독서에 영향을 미치는 변수로 나타났다(박소현, 2020).

마지막으로 기술환경적 요인은 최근 다양한 미디어의 등장과 관련이 있다. 책은 정보전달을 위한 미디어 중의 하나라는 점에서 다른 미디어와 경쟁 관계에 있다. 1일 가용 시간은 정해져있고 어떤 미디어를 선호하는가에 따라 학생들의 독서가 결정될 수 있다. 스마트폰은 전자책의 발달을 가속화했다는 긍정적 측면이 있지만 동시에 정보매체를 책에서 다른 많은 유형의 미디어로 다양화함으로써 정보 접근성을 확장하고 있다. 그러나 대부분의 연구에서는 스마트폰을 책과 대비시킴으로써 상호보완적이 아닌 이율배반적인 관계로 설정하였다. 스마트폰

사용 시간의 증가는 독서 시간의 감소를 의미하는 것으로 가정하고 연구를 진행했다(김태용, 박선주, 2014; 박소현, 2020; 이재선, 최선철, 2020). 그러나 스마트폰을 사용하는 시간이 문제라기보다는 스마트폰을 어떻게 사용하는지가 중요한 요인으로 나타났다. 독서 여부를 결정하는 요인으로 스마트폰 사용 목적, 즉 독서를 하는 학생은 스마트폰을 주로 정보 검색을 목적으로 사용한다는 점에서 차이를 보였다(박성재, 신인수, 2024). 부모의 미디어 사용 여부와 자녀의 독서시간과의 관계에 있어서도 유의미하지 않았고 오히려 자녀에게 미디어 사용 교육을 했는지에 대한 여부가 중요한 요인으로 나타났다(김아란 외, 2021).

2.2 전통적인 통계분석과 머신러닝 알고리즘 비교

전통적인 통계분석과 머신러닝 기법은 데이터 분석의 목적과 방법에서 차이가 있다. 가장 큰 차이는 전통적인 통계분석은 연역적으로 접근하고, 머신러닝은 귀납적으로 접근한다는 점이다. 전통적인 통계분석은 연구자가 사전에 설정한 가설을 검증하는 데 초점을 맞추고, 머신러닝 기법은 데이터로부터 패턴을 학습하여 예측 모델을 구축하는데 중점을 둔다. 두 접근법은 연구문제 설정부터 시작해서 연구 설계, 데이터 활용 방식, 결과 해석에 있어 차이가 있다.

전통적인 통계분석 방법(예: 회귀분석, 분산분석, 요인분석 등)은 독립 변수와 종속 변수 간의 관계를 명확하게 설명하는 데 주 목적이 있다. 연구자는 선행연구 분석 및 이론적 근거를 통해 특정 변수가 종속 변수(독서 여부 또는

독서 시간)에 미치는 영향을 검증하며, 가설검정을 통해 통계적으로 유의한 관계를 확인하려고 한다. 예를 들어, 선형 회귀 분석을 활용하면 부모의 학력, 가구 소득, 독서 태도 등이 독서 습관에 미치는 영향을 추정할 수 있고, 변수 간의 상호작용 효과도 검정할 수 있다. 그러나 전통적인 통계분석은 변수 선택이 연구자의 인식과 선행연구, 그동안 알려진 이론에 의존한다는 한계가 있다.

반면, 머신러닝 기법은 데이터의 분포나 관계를 사전에 가정하지 않고, 특정 알고리즘을 이용하여 대량의 데이터에서 패턴을 탐색하고 예측 모델을 최적화한다. 예를 들어, 의사결정나무(Decision Tree)나 랜덤 포레스트(Random Forest)와 같은 알고리즘은 독서 여부를 예측하는 데 있어 변수들의 상호작용을 고려하고, 변수 간 비선형적 관계도 포함한다. 전통적인 회귀 모델에서 가정하는 선형성 가정에 얽매이지 않는다. 또한, 머신러닝 모델은 독서 연구에서 비정형 데이터(예: 소셜미디어 게시물, 온라인 서평 등)도 분석할 수 있고, 예측 성능을 최대화하는 방향으로 모델을 최적화하는데 주 목적이 있다.

그러나 전통적인 통계분석은 가설검정과 분석 결과의 해석(예: 회귀 계수)을 통해 변수의 영향을 구체적으로 해석 및 설명할 수 있지만, 머신러닝 모델(특히 딥러닝 기반 모델)의 경우 특정 변수가 독서 여부에 미치는 영향을 명확하게 설명하기 어려울 수 있다. 이를 보완하기 위해 SHAP(SHapley Additive Explanations) 또는 PDP(Partial Dependence Plot)와 같은 기법이 활용되기도 한다. 또한, 머신러닝 모델은 적절한 하이퍼파라미터 튜닝과 데이터 전처

리가 필수적이며, 데이터 크기가 작거나 노이즈가 많은 경우 전통적인 통계분석보다 안정성이 떨어질 수 있다는 점을 유의해야 한다.

2.3 머신러닝 기법을 적용한 독서 연구

독서 연구 분야에도 머신러닝 기법이 적용되고 있다. 김영식 외(2019)는 머신러닝 기법 중의 하나인 랜덤 포레스트 기법을 적용하여 독서 활동 예측 요인을 분석했다. 고등학생들의 독서 활동 예측 요인으로 중학교 3학년 때의 성적, 혼자 공부하는 시간, 사교육 경험 여부, 활동 중인 동아리 개수 등이 영향력이 큰 것으로 나타났다. 특히, 이 연구에서는 머신러닝 기법의 특징 중의 하나인 탐색적 분석을 수행함으로써 기존 선행 연구에서 다루지 않은 변수들을 파악했다는 점에서 의미가 있다. 박찬홍(2022) 또한 랜덤 포레스트 기법을 적용하여 고등학생의 독서에 영향을 미치는 요인을 탐색하였다. 모델 정확도 개선 지수를 기준으로 요인을 순위화 한 결과, 주말 인터넷 및 TV 강의시간이 가장 중요한 변수로 나타났음을 밝히고 있다. 초·중학생을 대상으로 분석한 최지연(2024)의 연구 또한 랜덤 포레스트를 적용하여 독서 실행 영향 요인을 분석했다. 분석 결과, 중학교 1학년의 경우, 학습활동, 성취압력, 학교규칙, 교우관계 등의 학교 환경 변수가 영향을 미치는 것으로 보고하고 있다.

이들 연구는 랜덤 포레스트 기법을 주로 사용하여 영향 요인을 분석하고 있지만 이 외에도 다양한 머신러닝 기법이 있다는 점에서 기법 간 비교 연구가 필요하다. 또한, 머신러닝 기법을 활용하여 보다 다양한 변수와 복합적인

관계를 고려한 예측 모델을 구축함으로써 연구 결과의 현장성과 타당성을 높이하고자 한다.

3. 연구 설계

3.1 연구대상

본 연구에서는 청소년들의 독서 활동을 살펴 보기 위해 한국아동·청소년패널조사(KCYPS) 자료를 분석하였다. 해당 조사는 아동·청소년기 전반에 걸친 성장과 발달의 복합적인 변화를 체계적·다면적으로 파악하기 위해 시작되었으며, 2018년 기준 중학교 1학년 총 2,590명(남학생 1,450명, 여학생 1,294명)을 대상으로 매년 동일한 설문지를 활용해 자료를 수집하는 종단연구 형태로 진행되고 있다. 본 연구에서는 2021년 수집된 자료를 활용하여 분석을 수행하였다.

한국아동·청소년패널조사 데이터 응답자의 평일 독서시간을 조사한 결과는 <표 1>과 같다. 독서활동 '전혀 안함'에 해당하는 학생의 수는 677명으로 전체 2,590명의 26.1%에 해당하고, 나머지 '독서 활동 함'에 해당하는 학생의 수는 1,913명으로 73.9%에 해당한다. 독서집단과 비독서 집단을 구분하기 위해 비독서 집단은 '전혀 안함'이라고 응답한 677명(26.1%), 독서집단은 1시간 이상 독서를 한다고 한 472명(18.2%)으로 구분하였다.

3.2 측정도구

본 연구에서는 머신러닝 기법을 적용한 탐색

〈표 1〉 평일 독서시간별 학생수

집단 구분	독서시간	학생수	집단 구성원수	%	비고
비독서집단	전혀안함	677	677	26.1	
독서집단	30분미만	830	1,444	55.6	
	30분~1시간미만	611			
	1시간~2시간미만	314	472	18.2	
	2시간~3시간미만	86			
	3시간~4시간미만	36			
	4시간이상~	36			
합계		2,590	2,590	100.0	

적 데이터 분석을 위해 대규모 패널 데이터에서 분석에 부적합하거나 활용하기 어려운 변수들을 제거하거나 일부를 재코딩하였다. 먼저, 중1 코호트에 해당하는 자료 중 가구·개인 식별번호(ID), 코호트 구분, 학교코드, 가중치 등 분석에 직접 활용하기 어려운 항목을 제외했다. 또한 학교 및 거주지를 17개 시·도 단위로 응답받는 문항은 지역별 응답 분포가 크게 달라 더미 변수로 변환하기 부적합하다고 판단하여 삭제하였다. 추가적으로 평일·주말의 기상·취침 시각을 시·분 단위로 입력하는 문항이나 부모와 함께 보내는 시간(평일·주말 시·분)처럼 시간 정보를 세분화해 기입해야 하는 항목들은 분석 대상에서 제외하였다. 마지막으로, 일부 항목에서 ‘해당 없음’ 혹은 ‘잘 모르겠음’으로 답한 학생들은 결측치로 재분류하여 추가 처리하였다.

3.2.1 독립변수

변수 정리 과정을 거친 결과, 중1 코호트에서 종속변수를 제외하고 총 295개의 독립변수를 최종적으로 선정하여 분석에 반영하였다. 한국청소년정책연구원의 종단연구는 변수를 크게 개인발달 영역과 발달환경 영역으로 구분하

는데, 개인발달 영역은 생활시간, 지적발달, 사회·정서·역량발달, 신체발달, 비행 등으로, 발달환경 영역은 가정, 매체, 학교, 활동·문화 환경 영역으로 구성된다. 본 연구에서는 각 설문 문항별 정보를 최대한 보전하기 위해, 하나의 문항을 하나의 독립변수로 활용하였다. 또한 범주형으로 조사된 변수들은 더미 코딩을 통해 변환 후 분석에 포함하였다.

3.2.2 종속변수

본 연구에서 종속변수는 “주말 독서시간”으로 1 = ‘전혀 안함’, 2 = ‘30분 미만’, 3 = “30분 ~ 1시간 미만”, 4 = “1시간 ~ 2시간 미만”, 5 = “2시간 ~ 3시간 미만”, 6 = “3시간 ~ 4시간 미만”, 7 = “4시간 이상 ~”으로 측정된 변수를 더미 코딩으로 변환하여, ‘전혀 안함’을 0으로 하고, 1시간 이상한 집단을 ‘독서 활동 함’ 변수 1로 코딩 변경하여 사용하였다.

3.3 분석방법

본 연구는 청소년의 독서 여부를 예측하기 위해 다양한 머신러닝 기법을 적용하고, 각 기법의 예측 성능을 비교하여 가장 성능이 좋은

알고리즘과 독서 여부 결정요인을 귀납적으로 탐색하고자 한다. 이를 위해 데이터 전처리 및 표준화 과정을 수행한 후, 대표적인 머신러닝 기법들 중에서 로지스틱회귀, 의사결정나무, 랜덤 포레스트, adaptive LASSO, 서포트 벡터 머신, 그래디언트 부스팅, k-최근접 이웃 알고리즘을 적용하였다. 연구의 분석 절차는 다음과 같다.

3.3.1 데이터 전처리 및 데이터 변환(표준화)

데이터 분석을 수행하기 전에, 전처리 과정을 거쳤다. 한국아동·청소년패널조사 데이터를 기반으로 결측치 처리, 변수 변환, 표준화를 하였다. 첫째, 결측치 처리를 위해 30% 이상의 결측치를 가진 변수는 제거하였고, 30% 미만의 결측치는 다중 대체법(MICE, Multivariate Imputation by Chained Equations)을 사용하였다. 둘째, 범주형 변수는 원-핫 인코딩(One-Hot Encoding)을 통해 변환하였다. 셋째, SVM 및 신경망 모델의 학습 성능을 향상시키기 위해 연속형 변수를 평균 0, 표준편차 1로 변환하는 z-score 표준화를 수행하였다. 마지막으로, 데이터를 80:20 비율로 훈련(training) 데이터와 테스트(test) 데이터로 분할하였으며, 훈련 데이터 내에서 5-fold 교차 검증을 수행하여 최적의 하이퍼파라미터를 탐색하였다.

3.3.2 로지스틱 회귀(Logistic Regression)

로지스틱 회귀는 지도학습(Supervised Learning) 기반의 머신러닝 알고리즘으로, 종속 변수가 이분변수(binary)일 때 활용되는 분류 기법이다. 본 연구에서는 독서 여부(독서함=1, 독서안함=0)를 종속 변수로 설정하여 로지스틱 회귀

분석을 수행하였다. 시그모이드(Sigmoid) 함수를 사용하여 독립 변수와 종속 변수 간의 관계를 비선형적으로 변환하여 출력값이 0과 1 사이의 확률 값으로 해석할 수 있다. 학습된 모델에서 각 독립 변수의 회귀 계수(coefficients)는 해당 변수가 독서 여부에 미치는 영향을 나타내며 이는 모델의 해석 가능성을 높이는 장점이 있다. 본 연구에서는 R 패키지 glm을 사용하여 로지스틱 회귀 모델을 구축하였으며, 변수 간 다중공선성을 완화하기 위해 단계적 변수 선택(stepwise selection) 기법을 적용하여 최적의 변수 조합을 선정하였다.

3.3.3 의사결정나무(Decision Tree)

의사결정나무는 데이터를 트리(tree) 구조로 분류하여 예측하는 지도학습 알고리즘이다. 본 연구에서는 CART(Classification and Regression Tree) 알고리즘을 활용하여 노드를 분할하였다. 노드 분할 기준으로는 지니지수(Gini Index)를 적용하였으며, 과적합(overfitting) 방지를 위해 가지치기(pruning)를 수행하였다. 또한, 트리의 최대 깊이(max depth)와 최소 분할 샘플(min samples split)을 조정하여 최적의 모델을 선정하였다. R 패키지 rpart를 사용하여 모델을 구축하였다.

3.3.4 랜덤 포레스트(Random Forest)

랜덤 포레스트는 여러 개의 의사결정나무를 조합하여 예측 정확도를 향상시키는 앙상블(ensemble) 기법이다. 랜덤 포레스트는 개별 트리들이 서로 독립적인 학습을 수행하고, 최종 예측값을 다수결 방식(voting)으로 결정한다. 본 연구에서는 1,000개의 의사결정나무를 생성

하였으며, 부트스트래핑(Bootstrapping) 기법을 적용하였다. 또한, 나무의 마디 분할 시 랜덤으로 선택되는 변수 수(mtry)를 총 독립변수 개수의 제곱근으로 설정하였다. 모델 학습은 R 패키지 randomForest를 사용하여 수행되었다.

3.3.5 Adaptive LASSO

Adaptive LASSO(Adaptive Least Absolute Shrinkage and Selection Operator)는 전통적인 LASSO 회귀 분석을 개선한 방법으로, 가중치를 추가하여 변수 선택의 일관성을 높이는 기법이다. 이 방법은 다중공선성 문제를 해결하면서도 중요 변수를 선택하는 데 유용하다. 본 연구에서는 L1 정규화(벌점 회귀)를 적용하여 변수 선택을 수행하였으며 최적의 조율 모수(lambda)는 10-fold 교차 검증을 통해 결정하였다. 분석에는 R 패키지 glmnet을 사용하였다.

3.3.6 서포트 벡터 머신(Support Vector Machine, SVM)

SVM은 초평면을 이용하여 데이터를 분류하는 지도학습 기법이다. 본 연구에서는 비선형 데이터 분류를 수행하기 위해 RBF(Radial Basis Function) 커널을 적용하였다. 하이퍼파라미터 C(비용)와 gamma(감마값)는 5-fold 교차 검증을 통해 최적값을 선정하였으며, 분석에는 R 패키지 e1071을 사용하였다.

3.3.7 그래디언트 부스팅(Gradient Boosting)

그래디언트 부스팅은 손실 함수를 최소화하는 방향으로 여러 개의 결정 트리를 결합하는 앙상블 기법이다. 본 연구에서는 XGBoost

(eXtreme Gradient Boosting) 모델을 활용하여 분석을 수행하였다. 학습률(shrinkage)은 0.1, 나무 수(n.trees)는 500, 깊이(interaction depth)는 3으로 설정하였으며, 과적합 방지를 위해 정규화 파라미터를 적용하였다. 분석에는 R 패키지 gbm을 사용하였다.

3.3.8 k-최근접 이웃(k-Nearest Neighbors, kNN)

kNN은 새로운 데이터 포인트를 기존 데이터 포인트들과의 거리를 기반으로 분류하는 비모수적(non-parametric) 알고리즘이다. 본 연구에서는 독서 여부를 예측하기 위해 유클리드 거리(Euclidean distance)를 사용하였다. 최적의 k값을 선정하기 위해 5-fold 교차 검증을 수행하였으며, R 패키지 class를 사용하였다.

3.3.9 모델 평가 및 비교

각 모델의 성능을 비교하기 위해 예측정확도, 민감도, 특이도를 주요 평가지표로 사용하였다. 세 지표는 예측 모델이 양성(positive)과 음성(negative)을 얼마나 정확히 구분하는지 서로 다른 관점에서 보여준다. 예측정확도는 전체 데이터 중에서 모델이 올바르게 예측한 비율을 의미하고 공식 1을 이용하여 산출된다. 편향된 분포에서도 높게 나타날 수 있다는 점에서 정확도 외에 다른 지표도 함께 고려할 필요가 있다.

공식 1. 예측정확도(Accuracy) = $\frac{\text{정확하게 예측한 개수}}{\text{전체 개수}}$

민감도는 실제 양성 데이터 중에서 모델이

양성으로 정확히 예측한 비율을 의미한다. 민감도가 높을수록 실제 양성 데이터를 놓치지 않고 잘 찾아내는 모델이다. 특이도는 민감도와는 반대로 실제 음성 데이터 중에서 모델이 음성으로 정확히 예측한 비율이다. 특이도가 높다는 것은 실제 음성인 데이터를 양성으로 잘못 판단하지 않는 능력이 우수함을 의미한다. 이처럼 예측정확도, 민감도, 특이도를 함께 살펴보면 각 모델이 양성과 음성을 얼마나 잘 분류하는지를 종합적으로 판단할 수 있다. 본 연구에서는 모든 모델에 대해 세 가지 지표의 값을 구한 뒤, 가장 높은 성능을 나타낸 모델을 선정하였다.

공식 2. 민감도(Sensitivity) = 실제 양성 중 정확히 예측한 개수 / 전체 양성 개수

공식 3. 특이도(Specificity) = 실제 음성 중 정확히 예측한 개수 / 전체 음성 개수

3.3.10 주요 예측변수 결정 및 부분의존성 도표

가장 예측력이 높은 기법을 중심으로 주요 예측변수를 도출하고, 이러한 변수들이 독서 여부에 미치는 영향을 시각적으로 확인하기 위해 부분의존성 도표(Partial Dependence Plot, PDP)를 제시하였다. 머신러닝 기법을 적용하여 다양한 변수들의 상대적 중요도를 평가한 결과, 독서 여부를 결정하는 데 있어 유의미한 영향을 미치는 상위 10개의 예측변수를 선정하였다. 변수의 중요도는 특정 머신러닝 모델(예: 랜덤 포레스트, 그래디언트 부스팅 등)에서 제공하는 변수 중요도 지수(Feature Importance Score)를 기반으로 평가하여 선정하였다.

4. 분석결과

4.1 머신러닝 기법별 독서 예측 정확도 비교

청소년들의 독서활동을 예측하기 위한 머신러닝 기법으로 7가지 모델을 적용하였다. 각 모델별 예측정확도와 민감도, 특이도를 분석한 결과는 <표 2>와 같다. 예측정확도는 전체 데이터 중에서 모델이 맞춘 데이터의 수를 의미하고 전체 1,149명 중에 독서를 하는 학생은 472명, 독서를 하지 않는 학생은 677명으로 나타났는데 이들을 정확하게 분류한 비율을 의미한다. 예측정확도가 가장 높은 모델은 88%의 정확도를 보인 랜덤 포레스트 모형으로 나타났고 다음으로 그래디언트 부스팅(85%), SVM(83%), kNN(80%), 로지스틱 회귀(78%), adaptive LASSO(76%), 의사결정나무(74%) 순으로 나타났다.

실제로 양성(positive)인 사례를 놓치지 않고 잘 찾아내는 능력을 나타내는 민감도 분석결과에 있어서도 예측정확도와 유사한 결과를 보였다. 랜덤 포레스트 모델이 민감도가 87%로 가장 높게 나타났다. 즉, 랜덤 포레스트 모델은 독서를 하는 학생 472명 중에서 87%인 약 410명을 정확하게 예측하고 있다. 이와 반대로, 특이도는 음성(negative)인 사례를 잘 찾아내는 능력을 의미하고 본 연구에서는 독서를 하지 않는 학생을 정확하게 예측하는 정도를 의미한다. 특이도 분석 결과, 예측정확도와 민감도에 따른 모델 순위 결과와 동일하게 나타났다.

예측정확도와 민감도, 특이도를 종합적으로 고려했을 때, 7개의 모델 중에서 가장 좋은 성능

〈표 2〉 머신러닝 기법별 예측정확도, 민감도, 특이도 비교

모델	예측정확도	민감도	특이도	순위
랜덤 포레스트	0.88	0.87	0.89	1
그래디언트 부스팅	0.85	0.84	0.86	2
SVM	0.83	0.81	0.82	3
kNN	0.80	0.79	0.81	4
로지스틱 회귀	0.78	0.77	0.76	5
adaptive LASSO	0.76	0.75	0.74	6
의사결정나무	0.74	0.73	0.72	7

을 보인 모델은 랜덤 포레스트로 세 가지 영역 모두에서 가장 좋은 성능을 보였다. 다음으로 그래디언트 부스팅이 우수한 성능을 보이고 있었다. 랜덤 포레스트와 그래디언트 부스팅은 개별 모델을 여러 개 묶어 사용하는 앙상블 기법에 기반한다. 랜덤 포레스트는 의사결정나무를 무작위로 여러 개 생성하고 이들을 앙상블해 최종 예측을 수행하고 그래디언트 부스팅은 의사결정나무를 순차적으로 학습하면서 오차를 점진적으로 보완해 가는 방식을 취한다는 점에서 높은 예측정확도를 나타낸 것으로 보인다. 의사결정나무는 훈련 데이터에 지나치게 맞추어 학습하는 경향이 있으며, 새로운 데이터에 대한 일반화 성능이 낮아지는 문제가 발생할 수 있다.

4.2 독서 여부를 예측하는 주요 변인 탐색

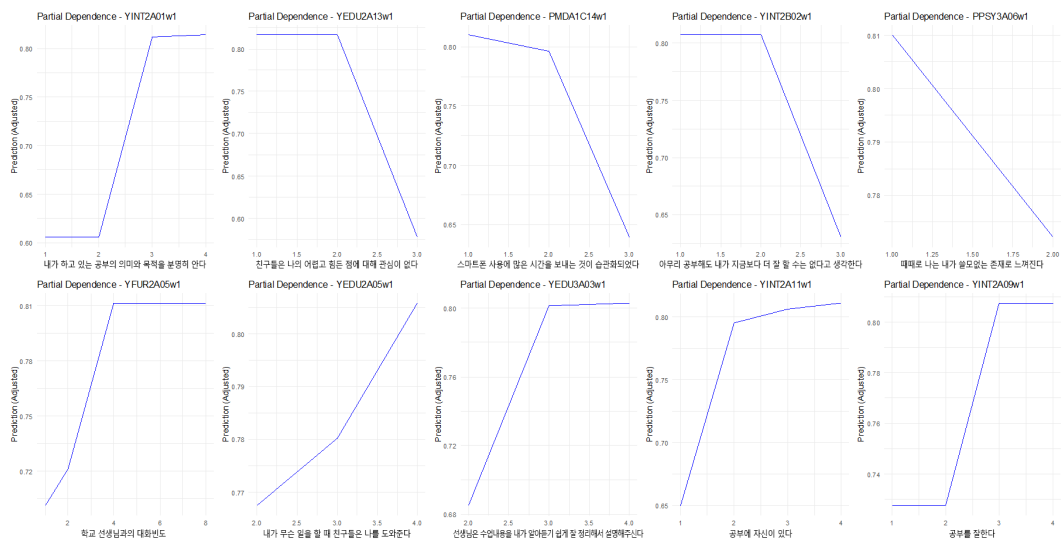
학업 열의와 관련된 질문 중 ‘내가 하고 있는 공부의 의미와 목적을 분명히 안다’가 가장 높은 중요도를 나타냈다. 중요도 변수 상위 10개 중에서 학업 열의의 관련 다른 변수로는 ‘공부에 자신이 있다’, ‘공부를 잘한다’ 등이 포함되어 있지만 순위로는 9위와 10위를 차지하고 있다. 다음으로 친구관계에 대한 질문 중 ‘친구들은 나

의 어렵고 힘든 점에 대해 관심이 없다’라는 다소 부정적인 요인이 독서 여부를 결정하는 중요한 변수로 나타났다. 또한 ‘스마트폰 사용에 많은 시간을 보내는 것이 습관화되었다’나 ‘아무리 공부해도 내가 지금보다 더 잘할 수는 없다고 생각한다’, ‘때때로 나는 내가 쓸모없는 존재로 느껴진다’는 부정적 습관과 정서를 나타내는 변수가 중요도에서 높은 순위를 차지하고 있었다. 교사와의 관계를 나타내는 진로 관련 대화 빈도와 ‘내가 무슨 일을 할 때 친구들은 나를 도와준다’와 같은 친구관계 설문, ‘선생님은 수업내용을 내가 알아듣기 쉽게 잘 정리해서 설명해주신다’와 같은 교사관계가 중요도 상위 10개 변수에 포함되었다.

〈표 3〉에서 제시된 상위 10개의 변수는 독서하는 학생과 독서를 하지 않는 학생을 구분하지 않고 예측하는데 중요한 변수의 순위를 바탕으로 했다는 점에서 두 변수간 관계의 내용을 분석하기 위해서 부분의존성 도표를 확인했다. 부분의존성 도표의 X축은 예측변수 값의 범위를 나타내고 Y축은 종속변수의 추정값이다. 예측변수 값이 증가할수록 종속변수의 추정값, 즉 독서여부가 결정되는 것을 알 수 있다. 〈그림 1〉에 제시된 부분의존성 도표 중 1순위와 6순위, 7순위, 8순위, 9순위, 10순위 변수는

〈표 3〉 독서 예측 중요도 상위 10개 변수

순위	항목명 (변수명)	질문	중요도
1	학업열의 YINT2A01w1	내가 하고 있는 공부의 의미와 목적을 분명히 안다.	3.88
2	친구관계 YEDU2A13w1	친구들은 나의 어렵고 힘든 점에 대해 관심이 없다.	3.79
3	스마트폰의존도 PMDA1C14w1	스마트폰 사용에 많은 시간을 보내는 것이 습관화되었다.	2.91
4	학업무기력 YINT2B02w1	아무리 공부해도 내가 지금보다 더 잘할 수는 없다고 생각한다.	2.51
5	자아존중감 PPSY3A06w1	때때로 나는 내가 쓸모없는 존재로 느껴진다.	2.37
6	진로관련대화빈도 YFUR2A05w1	학교 선생님과과의 대화빈도	2.36
7	친구관계 YEDU2A05w1	내가 무슨 일을 할 때 친구들은 나를 도와준다.	2.19
8	교사관계 YEDU3A03w1	선생님은 수업내용을 내가 알아듣기 쉽게 잘 정리해서 설명해주신다.	2.15
9	학업열의 YINT2A11w1	공부에 자신이 있다.	2.14
10	학업열의 YINT2A09w1	공부를 잘한다.	2.10



〈그림 1〉 주요 예측변수의 부분의존성 도표

값이 증가할수록 독서할 확률이 증가한다. 공부의 의미와 목적을 알고 있는 등의 학업열의가 높을수록, 선생님과 진로 관련 대화의 빈도가 많을수록, 친구 관계가 좋을수록, 교사관계가 좋을수록, 공부 자신감이나 자신이 공부를 잘한다고 생각할수록 학생들의 독서할 확률은 상승한다. 이와 반대로 2순위부터 5순위 변수는 값이 증가할수록 독서할 확률이 감소하는 것을 알 수 있다. 친구의 무관심이나 스마트폰 의존, 학업 무기력, 낮은 자아존중감은 독서할 확률 감소에 영향을 미치는 것으로 나타났다.

또한 부분의존성 도표는 예측변수가 어떤 구간에서 독서 여부가 바뀌는지를 보여주고 있다. 학업열의 관련 변수 중에서 공부의 의미와 목적을 알고 있는지와 공부를 잘한다는 질문에 ‘매우 그렇지 않다’와 ‘그렇지 않다’를 선택한 집단은 독서를 하지 않은 집단으로, ‘그렇다’와 ‘매우 그렇다’를 선택한 집단은 독서를 하는 집단으로 분류된다. 학업열의 관련 변수 중에서 공부 자신감에 대한 질문에 ‘그렇지 않다’고 응답한 경우에도 독서를 하는 집단으로 분류가 되는 것으로 나타났다. 교사와의 관계에서 교사가 수업내용을 잘 설명해주는지에 대한 응답에 있어서도 ‘그렇다’와 ‘매우 그렇다’를 선택한 학생들이 독서집단으로 분류될 확률이 높았다. 교사와의 진로 관련 대화 빈도가 증가할수록 독서할 확률이 상승하지만 4회 이상에서는 변화가 거의 없는 것으로 나타났다. 이들 변수들은 모두 일정 수준 이상에서 포화 상태를 보인 반면에 친구관계를 나타내는 변수에 있어서는 선형 관계를 보였다. 즉, 자신이 무슨 일이 있을 때 친구들이 자신을 도와준다는 긍정적인 응답

이 높아질수록 독서 집단에 포함될 확률이 상승했다.

문항에 대한 응답과 독서확률이 음의 관계를 보이는 변수 중 친구들의 무관심, 스마트폰 의존도, 학업 무기력은 ‘매우 그렇지 않다’와 ‘그렇지 않다’ 응답에서 독서 집단에 포함될 확률이 높았지만 ‘그렇다’ 응답에서는 비독서 집단으로 분류되었다. 자아존중감이 감소할수록 독서집단에 포함될 확률이 감소했다. 두 변수 간의 관계는 선형성을 띠는 점에서 앞선 세 변수(친구들의 무관심, 스마트폰 의존, 학업무기력)와는 다른 형태를 보였다.

5. 논 의

5.1 독서 여부 예측을 위한 머신러닝 기법 비교

데이터의 양이 증가하고 분석 기술이 발전함에 따라 다양한 머신러닝 기법들이 제안되었다. 본 연구에서는 머신러닝 기법 중에서 연구에서 빈번히 언급되는 7개의 모델을 대상으로 독서 여부 예측 성능을 비교했다. 분석 결과, 랜덤 포레스트 모델이 예측정확도, 민감도, 특이도 등 모든 측면에서 가장 성능이 좋은 모델로 나타났다. 다음으로는 그래디언트 부스팅 모델이 상대적으로 우수한 성능을 보였다. 이 두 모델은 앙상블 기법을 적용했다는 점에서 다른 모델들보다 우수한 성능을 보인 것으로 판단된다. 주요 예측변수의 부분의존성 도표에도 나타났듯이, 변수들은 선형관계로 나타나지 않는다는 점에서 비선형 관계와 변수 간 상호작용을 효과적으

로 포착할 수 있는 랜덤 포레스트와 그래디언트 부스팅 기법이 독서 예측에서 높은 성능을 나타냈다. 청소년 독서 여부는 다양한 심리적, 사회적, 환경적 요인의 상호작용에 의해 결정되므로, 독립변수와 종속변수 간의 관계가 선형적이지 않을 가능성이 크다고 할 수 있다.

인적자본기업패널 데이터를 활용하여 직무 만족과 조직몰입에 영향을 미치는 요인을 탐색한 박소영, 정혜원(2019)은 LASSO, Ridge, 랜덤 포레스트, SVM 모델의 예측 성능을 비교했다. 네 가지 모델 중에서 랜덤 포레스트 모델이 직무만족과 조직몰입을 예측하는 정확도가 가장 높았다. 사교육 참여 예측을 위해 랜덤 포레스트, 나이브 베이즈 분류, SVM, 인공신경망 모델의 적합도를 탐색한 김영식, 김훈호(2019)의 연구 결과 또한 랜덤 포레스트 모델이 상대적으로 좋은 성능을 보였다. 앞서 사용한 모델(박소영, 정혜원, 2019) 이외에 추가로 모델을 투입하여 6가지 모델을 비교한 박소영, 정혜원(2021)은 한국아동청소년패널 데이터를 활용하여 중학생의 진로 결정 예측변수를 탐색하였다. 연구 결과, 가장 좋은 성능을 보인 모델은 그래디언트 부스팅이었고 다음으로 adaptive LASSO, 랜덤 포레스트 모델 순으로 나타났다. 그러나 이와는 다른 연구 결과도 있는데 이러닝 환경에서 학업성취에 영향을 미치는 요인을 머신러닝 기법을 적용하여 연구한 조한국(2018)의 연구에서는 SVM이 가장 좋은 성능을 보였고 그래디언트 부스팅이 가장 성능이 좋지 않은 것으로 나타났다. 따라서 다양한 데이터 세트를 활용하여 머신러닝 기법 간 성능차이가 어느 요인에서 기인하는지를 종합적으로 비교하고 검증할 필요가 있다.

5.2 학업열의와 독서

연구결과, 학업열의와 관련된 세 개의 문항이 독서여부에 영향을 미치는 중요한 변수 상위 10개에 포함되어 있다. 특히, '내가 하고 있는 공부의 의미와 목적을 분명히 안다'는 문항은 중요도가 3.88로 가장 높게 나타났고 '공부에 자신이 있다', '공부를 잘한다' 문항 또한 각각 2.14, 2.10의 중요도를 보였다. 부분의존성 도표에서 이들 변수는 독서여부와 정적인 관계를 보였다. 이는 공부의 의미와 목적을 분명히 알고 공부에 대한 자신감과 잘 한다는 인식이 높은 학생이 독서를 할 확률이 높음을 의미한다. 중학생의 독서시간과 학업열의 관계를 연구한 이은주 외(2024)의 연구 결과 또한 독서시간과 학업열의 간의 정적인 관계를 밝히고 있다. 학업열의와 함께 함께 논의할 수 있는 항목으로 상위 10개 변수에 포함된 학업무기력이 있다. 학업무기력은 학업열의와 반대되는 개념으로 사용되었으며 '아무리 공부해도 지금보다 더 잘할 수는 없다고 생각한다'는 문항이 2.51의 높은 중요도를 나타냈다. 부분의존성 도표를 통해 해당 변수는 독서여부와 부적인 관계를 나타냈다는 점에서 학업무기력이 높을수록 독서를 하지 않는 경우가 많은 것으로 나타났다. 독서집단과 비독서집단을 구분하는 요인에 대해 연구한 박성재, 신인수(2024)의 연구 또한 학업무기력이 비독서에 영향을 미치는 것으로 보고하고 있다. 다른 관점으로 독서가 학업무기력을 감소시킬 수 있다는 연구결과(김나영, 황혜영, 2020)는 두 변수 간의 관계뿐만 아니라 학업무기력 문제를 해결하고 학업열의를 끌어올릴 수 있는 정책, 즉 독서를 통해 학업무

기력을 학업열의로 바꾸는 정책 마련이 필요할 것이다.

5.3 친구 및 교사관계와 독서

중학생 응답자들에게 있어 친구와 교사 관계가 독서에 영향을 미치는 것으로 나타났다. 관계 관련 변인 중에서 친구들의 무관심 변수의 중요도가 3.79로 전체 변수 중에서 두 번째로 높은 중요도를 나타냈다. 부분의존성 도표에서 친구들의 무관심은 부적인 관계를 나타냈다는 점은 친구들이 자신에게 관심이 없다는 느끼는 학생들은 비독서 집단으로 분류되는 것을 의미한다. 더욱 의미가 있는 분석은 무관심 변수가 친구들의 관심 변수의 중요도(2.19)보다 높게 나타났다는 점에서 오히려 무관심 문제의 해결이 우선시되어야 한다. 특히, 교사나 부모와의 관계보다 친구와의 관계가 독서에 미치는 영향이 크다는 점(정대근, 박성우, 2015)에서 친구와의 독서 관계 구축을 통해 무관심을 관심으로 전환할 필요가 있다.

연구 대상으로 하는 중학생들의 사회적 관계가 주로 학교 교육 현장을 바탕으로 하고 있다는 점에서 교사와의 관계 또한 독서에 중요한 영향을 미친다. 본 연구에서도 교사와의 관계 관련 변수의 중요도가 2.15로 나타났고 상위 10개 변수에 포함되었다. 이외에도 진로 관련 교사와의 대화 빈도 또한 중요도 2.36으로 나타났다는 점에서 독서여부에 있어서 교사와의 관계의 중요성을 확인할 수 있다. 본 연구에서 교사와의 관계를 나타내는 문항은 '선생님은 수업 내용을 내가 알아듣기 쉽게 잘 정리해서 설명해준다'로 독서와 관련성이 높지는 않지만 독

서에 있어서도 교사의 역할모델이 중요하다는 점(Gambrell, 1996)에서 독서를 통한 학생과 교사와의 관계 설정 또한 필요할 것이다.

5.4 스마트폰 의존과 독서

독서여부에 영향을 미치는 요인 중의 하나는 스마트폰 의존도로 나타났다. 스마트폰 의존도는 2.91의 중요도를 나타냈고 전체 순위에서 세 번째로 중요한 변수이다. 스마트폰 의존도에 대한 구체적인 문항으로 많은 시간을 보내는 습관이 높게 나타났는데 이는 부분의존성 도표에서 독서여부와 부적인 관계를 보였다. 스마트폰 의존도가 높은 학생은 독서를 하지 않고 의존도가 낮은 학생일수록 독서를 하는 것으로 해석할 수 있다. 스마트폰 의존과 독서시간은 다른 연구에서도 그 관련성이 입증되었는데 김태용, 박선주(2014)의 연구에서 스마트폰 보유여부가 독서시간과 독서량에 영향을 미치는 것으로 나타났다. 또한 이재선, 최선철(2020)의 연구에서도 스마트폰 사용시간이 높을수록 독서시간이 줄어드는 것을 발견했다. 독서여부에 영향을 미치는 요인에 대한 탐색적 연구를 수행한 박성재, 신인수(2024)의 연구에서도 독서시간과 스마트폰 사용시간은 부적인 관계를 나타내고 있다. 다만, 그들의 연구에서는 스마트폰 사용 시간과 함께 사용의 목적에 따라 독서여부가 결정되는 것을 보고하고 있는데 스마트폰을 주로 정보검색에 이용하는 학생들은 독서를 하는 경향이 나타났고 스마트폰을 TV나 동영상 시청을 목적으로 사용하는 학생들은 비독서 집단으로 분류되는 경향이 나타났다.

이러한 결과를 바탕으로 스마트폰 사용에 대

해 지도하거나 정보검색 등 생산적인 활용을 독려한다면 독서 활동에 긍정적인 효과를 기대할 수 있을 것이다. 또한 스마트폰 사용을 제한하기 보다는 독서 앱이나 디지털 독서 콘텐츠를 적극 활용하도록 장려하는 접근을 통해 독서 태도를 개선하고 청소년들이 주도적으로 읽기 활동을 지속하게 할 수 있는 정책이 마련된다면 청소년들의 독서율 상승과 학습 동기 제고를 가져올 것으로 기대된다.

5.5 자아존중감과 독서

자아존중감 관련 문항 중 '때때로 나는 내가 쓸모없는 존재로 느껴진다'는 문항은 독서여부를 결정하는데 중요한 변수로 나타났다. 전체 5순위로 그 중요도는 2.37이다. 부분의존성 도표에서 자아존중감 변수는 독서여부와 부적인 관계를 나타냈다. 자아존중감이 낮을수록 독서를 하지 않고 자아존중감이 높을수록 독서를 한다는 의미이다. 자아존중감과 독서의 관계는 독서치료 프로그램의 효과를 측정하는 연구의 하나로 진행되었고 독서가 자아존중감 향상에 긍정적인 영향을 미치고 있음을 보고하고 있다(강선숙, 오의경, 2016; 김갑임, 2008; 배현징, 2010; 서나라, 2006; 서순영, 2007; 이영림, 2007; 조현경, 손현동, 2015).

그러나 이은주 외(2024)의 연구에서는 독서 시간이 자아존중감에 유의미한 영향을 미치지 않는 것을 보고하고 있다. 또한, 박주현 외(2015)는 자아존중감을 4개의 하위개념(총체적, 사회적, 가정, 학교)으로 구분하고 책 읽어주기 프로그램이 초등학생의 자아존중감에 어떤 영향을 미치는지를 연구했는데 그 결과, 사회적 자

아존중감에만 통계적으로 유의미한 변화가 있었을 뿐 다른 영역에서는 유의미한 변화가 나타나지 않았다. 지역아동센터 아동을 대상으로 한 박길여(2011)의 연구에서는 책 읽어주기 활동이 인지적, 신체적 자아존중감 향상을 가져왔지만 사회적 자아존중감에는 영향을 미치지 않는다고 보고하고 있다. 이러한 연구 결과의 차이는 향후 연구에서 독서와 자아존중감의 관계를 다시 밝힐 필요가 있음을 보여준다. 또한 그 관계에 있어서 대상과 프로그램의 내용 등 연구 상황에 따라 달라질 수 있다는 점에서 다양한 상황에서의 연구와 분석이 요구된다.

6. 결론 및 연구 제한점, 후속 연구 과제

본 연구는 중학생을 대상으로 독서 여부를 예측하기 위하여, 머신러닝 기법 일곱 가지를 적용하고 그 성능을 비교하였다. 분석 결과, 앙상블 기법인 랜덤 포레스트와 그레디언트 부스팅이 다른 모델(SVM, kNN, 로지스틱 회귀, adaptive LASSO, 의사결정나무)에 비해 더 높은 예측 정확도와 민감도, 특이도를 보였다. 특히 랜덤 포레스트가 종합 성능(Accuracy, Sensitivity, Specificity)에서 1위를 차지하여 비선형적이고 복잡한 독서 결정 요인을 포착하는 데 유리함이 확인되었다. 또한 모델을 통해 도출한 변수 중요도와 부분의존성 도표 분석에서, '학업 열의', '친구·교사 관계', '스마트폰 의존도' 등이 중학생의 독서 여부를 결정하는 핵심 요인으로 나타났다. 구체적으로 학업 열의("공부에 자신 있음", "공부 의미를 잘 인

지”)가 높을수록 독서 확률이 높아지는 경향이 관찰되었다. 반면 스마트폰 의존, 학업 무기력, 부정적 자기인식 등은 독서율을 떨어뜨리는 주요 변인으로 나타났다.

이러한 결과는 기존 연구에서 제시된 독서와 학업성취 및 자기주도학습 간의 연관성 및 ‘스마트폰 사용시간 증가가 독서시간을 대체한다’는 가정과 일맥상통한다. 다만 본 연구는 머신러닝 기반의 분석을 통해, 기존 통계분석으로 확인하기 어려운 비선형적·계단형 패턴을 구체적으로 제시했다는 점에서 의의가 크다. 예컨대, 교사와의 진로 대화 빈도가 일정 수준까지 독서확률이 상승을 견인하지만 그 이후 포화되는 패턴이 발견되었다. 친구관계에서 ‘무관심’을 느끼는 정도와 독서여부 간에는 선형적 역관계가 뚜렷이 나타났는데, 친구 관심이 약해질수록 독서를 하지 않을 가능성이 큰 것으로 추정된다. 요약하면, 중학생들의 독서 여부는 학업과 자기효능감 같은 학업열의 요인, 교사 및 친구의 사회적 지원 요인, 그리고 스마트폰 의존도 등 새로운 미디어 환경 요인이 복합적으로 작용함을 확인할 수 있었다. 교육 현장에서 독서 장려 정책을 수립할 때, 단순히 도서 제공이나 독서 캠페인만으로는 한계가 있음을 시사한다. 오히려 학업 동기 부여(열의·자신감 증진), 우호적 교우·교사 관계 형성, 미디어 활용 지도를 결합한 입체적 지원이 필요하다고 볼 수 있다. 특히, 스마트폰에 대한 과도한 의존을 줄이고, 친구나 교사와의 교류를 통해 독서를 자연스럽게 권장하는 문화 형성이 효과

적일 것으로 예상된다. 결론적으로, 본 연구는 머신러닝 기법을 활용한 독서 예측 모델을 제시하고, 중학생 독서 행동의 주요 요인을 과학적·정량적으로 규명했다는 점에서 연구적·실천적 의의가 있다.

이러한 의의에도 불구하고 본 연구는 다음과 같은 제한점을 가진다. 첫째, 본 연구는 독서 여부만을 분석하였으며 독서의 내용이나 목적과 같은 질적 측면을 고려하지 못했다. 따라서 향후 연구에서는 독서의 내용과 유형별 차이를 분석하여 독서 행태를 보다 정교하게 이해할 필요가 있다. 둘째, 머신러닝 모델의 해석 가능성을 높이기 위해 부분의존성 도표를 활용하여 주요 변수들이 독서 여부에 미치는 영향을 분석하였다. 그러나 부분의존성 도표는 개별 변수와 독서 여부 간의 관계만을 단변량 분석(univariate analysis)으로 보여주기 때문에, 변수 간의 상호작용 효과를 설명하는 데 한계가 있다. 또한, 상위 10개 변수의 선정이 전체 데이터의 대표성을 충분히 반영하는지에 대한 검토가 부족할 수 있다. 따라서 후속 연구에서는 SHAP(SHapley Additive Explanations)이나 ICE(Individual Conditional Expectation) 기법을 활용하여 변수 간 상호작용을 고려한 보다 정교한 분석이 필요하다. 이러한 후속 연구를 통해 청소년의 독서 행태를 보다 심층적으로 이해하고 효과적인 독서 교육 및 장려 정책 수립에 필요한 근거를 제공할 수 있을 것이다.

참 고 문 헌

- 강선숙, 오의경 (2016). 독서교육프로그램의 자아존중감 및 독서태도에 대한 영향. 한국비블리아학회지, 27(3), 235-253. <http://doi.org/10.14699/kbiblia.2016.27.3.235>
- 곽길여 (2011). 지역아동센터에서 교사의 책읽어주기가 아동의 자아존중감과 독서태도에 미치는 영향. 석사학위논문, 순천대학교 대학원.
- 김갑임 (2008). 독서치료가 초등학생의 자아존중감과 학교생활 적응에 미치는 영향. 석사학위논문, 아주대학교 교육대학원.
- 김나영, 황혜영 (2020). Lasso 별점회귀모형을 적용한 중학생의 학업열의 관련 변수 탐색. 교육문화연구, 26(6), 337-355. <http://doi.org/10.24159/joec.2020.26.6.337>
- 김아란, 남민지, 최영은 (2021). 초등 4-6학년 아동의 미디어 사용, 실행기능과 독서 시간 및 선호도와와의 관계 탐색. 한국심리학회지: 발달, 34(4), 109-131. <http://doi.org/10.35574/KJDP.2021.12.34.4.109>
- 김영식, 김민석, 이길재 (2019). 랜덤 포레스트를 활용한 고교생의 독서활동 예측 요인 분석. 교육재정경제연구, 28(4), 137-156. <http://doi.org/10.46967/jefe.2019.28.4.137>
- 김영식, 김훈호 (2019). 머신러닝 기법을 활용한 사교육 참여 예측 모형 탐색. 교육재정경제연구, 28(3), 29-52. <http://doi.org/10.46967/jefe.2019.28.3.29>
- 김태용, 박선주 (2014). 초등학생의 스마트폰 사용 실태가 독서 실태 및 자기조절읽기에 미치는 영향. 정보교육학회논문지, 18(3), 433-442. <http://doi.org/10.14352/jkaie.2014.18.3.433>
- 남수정 (2023). 고등학생의 독서활동 및 부모-자녀 간 의사소통과 학업성취도와의 관계에서 학업 효능감의 매개효과. 석사학위논문, 부경대학교.
- 문화체육관광부 (2023). 2023년 국민 독서실태 조사.
- 박성재 (2022). 독서가 자기효능감과 다문화수용성에 미치는 영향 분석: 다변량 잠재성장모형을 적용하여. 한국도서관·정보학회지, 53(4), 293-318. <http://doi.org/10.16981/kliss.53.4.202212.293>
- 박성재, 신인수 (2024). 생성형 인공지능 시대 청소년의 독서에 영향을 미치는 요인 탐색. 정보관리학회지, 41(3), 171-203. <http://doi.org/10.3743/KOSIM.2024.41.3.171>
- 박소영, 정혜원 (2019). 머신러닝 기법을 적용한 기업 구성원의 직무만족 및 조직몰입에 영향을 미치는 요인 탐색. 교육학연구, 57(4), 183-213. <http://doi.org/10.30916/KERA.57.4.183>
- 박소영, 정혜원 (2021). 랜덤 포레스트를 활용한 고등학생의 진로개발역량 예측변수 탐색. 열린교육연구, 29(1), 239-265. <http://doi.org/10.18230/tjye.2021.29.1.239>
- 박소현 (2020). 초등학생과 중학생의 여가시간을 예측하는 변인들. 석사학위논문, 서울교육대학교 교육전문대학원.

- 박주현, 이명규, 박경희 (2015). 책 읽어주기가 초등학생들의 독서태도와 자아존중감에 미치는 영향. 한국도서관·정보학회지, 46(2), 1-27. <http://doi.org/10.16981/kliss.46.2.201506.1>
- 박찬홍 (2022). 초등학생 독서 시간의 종단적 변화가 중학교 1학년의 진로성숙도 및 자아개념에 미치는 영향. 한국초등국어교육, 74, 123-142.
- 박혜숙 (2008). 시간에 따른 중학생의 독서태도, 독서습관, 독서시간과 국어성취도와의 관계. 교육과학연구, 39(3), 27-49.
- 배현정 (2010). 독서치료가 초등학생의 자아존중감과 대인관계에 미치는 효과. 석사학위논문, 제주대학교 교육대학원.
- 변우열 (2012). 독서태도 모형 분석을 통한 독서태도 조사 설문 내용 개발. 한국도서관·정보학회지, 43(4), 139-159. <http://doi.org/10.16981/kliss.43.4.201212.139>
- 서나라 (2006). 독서치료 프로그램이 초등학교 저학년 아동의 자아존중감 및 학교생활 적응에 미치는 영향. 석사학위논문, 이화여자대학교 교육대학원.
- 서순영 (2007). 독서지도 프로그램이 아동의 자아존중감과 대인문제해결력에 미치는 영향. 석사학위논문, 진주교육대학교 교육대학원.
- 왕효성 (2014). 청소년 독자들의 독서 태도에 영향을 주는 요인에 관한 연구: 청소년 독자들이 말하는 책 읽기 '좋고 싫음'의 이유를 바탕으로. 독서연구, 33, 171-198. <http://doi.org/10.17095/JRR.2014.33.5>
- 이경남 (2016). 독해를 방해하는 요인분석 연구. 독서연구, 40, 119-147. <http://doi.org/10.17095/JRR.2016.40.4>
- 이국희 (2022). 생산적 독서 습관과 독서에 대한 해석이 독서에 미치는 효과. 한국콘텐츠학회논문지, 22(5), 565-575. <http://doi.org/10.5392/JKCA.2022.22.05.565>
- 이국희, 김은미, 최인철, 한다혜 (2020). 청소년의 독서 선호도와 독서에 대한 해석수준의 상호작용이 지속적인 독서와 독서량에 미치는 효과. 한국사회과학연구, 39(1), 5-37. <http://doi.org/10.18284/jss.2020.04.39.1.5>
- 이병기 (2015). 자기결정성 동기이론에 의한 독서동기 측정도구 개발에 관한 연구. 한국문헌정보학회지, 49(3), 311-332. <http://doi.org/10.4275/KSLIS.2015.49.3.311>
- 이영림 (2007). 독서치료프로그램이 청소년의 자아존중감에 미치는 영향. 석사학위논문, 백석대학교 사회복지대학원.
- 이은주, 김민정, 송연주 (2024). 중학생의 독서시간과 사교육시간이 학업열의에 미치는 영향에서 자아존중감의 매개효과. 교육종합연구, 22(2), 147-165. <http://doi.org/10.31352/JER.22.2.147>
- 이재선, 최선철 (2020). 스마트폰 사용 시간과 독서 시간 간의 관계에서 그릿(Grit)의 매개효과. 학습자중심교과교육연구, 20(7), 1009-1035. <http://doi.org/10.22251/jlcci.2020.20.7.1009>
- 정대근, 박성우 (2015). 독서역할모델이 고등학생의 독서유효성에 미치는 영향 연구. 한국도서관·정보

- 학회지, 46(3), 201-224. <http://doi.org/10.16981/kliss.46.3.201509.201>
- 정혜승 (2006). 읽기 태도의 개념과 성격. 독서연구, 16, 383-405.
- 조증열, 이임숙 (2004). 초등학생의 읽기 능력, 독서, 학업효능감 및 학업성적간의 인과적 관계. 초등교육연구, 17(1), 127-144.
- 조한국 (2018). 머신 러닝을 활용한 이러닝 학습 환경에서의 학습자 성취 예측 모형 탐색. 학습자중심교과교육연구, 18(21), 553-572. <http://doi.org/10.22251/jlcci.2018.18.21.553>
- 조현경, 손현동 (2015). 2009 개정 교육과정 국어교과서 작품을 활용한 독서치료가 초등학생의 자아존중감 향상에 미치는 효과. 교육치료연구, 7(3), 375-392.
- 최지연 (2024). 랜덤포레스트를 활용한 초·중학생 독서실행 영향 요인 탐색. 학습자중심교과교육연구, 24(8), 239-250.
- 하문선 (2017). 청소년의 생활시간 사용 중단적 변화와 가정환경, 정서행동문제 간 관계. 청소년학연구, 24(3), 25-59. <http://doi.org/10.21509/KJYS.2017.03.24.3.25>
- Baker, L. & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. Reading Research Quarterly, 3, 452-477. <https://doi.org/10.1598/RRQ.34.4.4>
- Gallagher, K. (2023). Readicide: How Schools are Killing Reading and What You can Do About It. New York: Routledge. <https://doi.org/10.4324/9781032682198>
- Gambrell, L. B. (1996). Creating classroom cultures that foster reading motivation. The Reading Teacher, 50(1), 14-25.
- Myers, J. (1992). The social contexts of school and personal literacy. Reading Research Quarterly, 27, 297-332. <https://doi.org/10.2307/747672>
- Schallert, D. L. & Reed, J. H. (1997). The Pull of the Text and the Process of Involvement in Reading. In Guthrie & Wigfield eds. Reading Engagement: Motivating Readers Through Integrated Instruction. Delaware: International Reading Associations, 68-85.
- Wigfield, A. & Guthrie, J. T. (1995). Dimensions of Children's Motivations for Reading: An initial Study (Research Report No. 34). Georgia: National Reading Research Center.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

- Bae, Hyun-Jeong (2010). The Effect of Bibliotherapy on Self-Esteem and Interpersonal Relationships of Elementary School Students. Master's thesis, Cheju National University.
- Byun, Woo-Yeol (2012). A development of the contents for the reading attitude survey questionnaire

- through the analysis of reading attitude models. *Journal of Korean Library and Information Science Society*, 43(4), 139-159. <http://doi.org/10.16981/kliss.43.4.201212.139>
- Cho, Hyun-Kyung & Son, Hyun-dong (2015). The effect of bibliotherapy utilizing Korean language textbook of 2009 educational curriculum revision on the self-esteem of elementary school students. *Korean Journal of Educational Therapist*, 7(3), 375-392.
- Cho, Jeung-Ryeul & Lee, Imsook (2004). Causal relations among reading ability, reading activity, academic self-efficacy and academic achievement of primary school children. *The Journal of Elementary Education*, 17(1), 127-144.
- Choi, Jiyoun (2024). Exploring influential factors on elementary and middle school students' reading practice using random forests. *The Journal of Learner-Centered Curriculum and Instruction*, 24(8), 239-250.
- Chung, Hyeseung (2006). The concept and character of an attitude toward reading. *Journal of Reading Research*, 16, 381-403.
- Ha, Moon-sun (2017). The relationship between the trajectory of a life time of adolescents, family environment and emotional behavioral problems. *Korean Journal of Youth Studies*, 24(3), 25-59. <http://doi.org/10.21509/KJYS.2017.03.24.3.25>
- Jeong, Dae Keun & Park, Seong Woo (2015). A study on impact of reading role model influencing reading effectiveness to high students. *Journal of Korean Library and Information Science Society*, 46(3), 201-224. <http://doi.org/10.16981/kliss.46.3.201509.201>
- Jho, Hunkoog (2018). Exploration of predictive model for learning outcomes of students in the e-learning environment by using machine learning. *The Journal of Learner-Centered Curriculum and Instruction*, 18(21), 553-572. <http://doi.org/10.22251/jlcci.2018.18.21.553>
- Kang, Sun-Sook & Oh, Eui-Kyung (2016). A study on the effect of reading education program on self-esteem and reading attitude: The case of M high school students. *Journal of the Korean Biblia Society for Library and Information Science*, 27(3), 235-253. <http://doi.org/10.14699/kbiblia.2016.27.3.235>
- Kim, Aran, Nam, Minji, & Choi, Youngon (2021). Exploring the relationship between digital device usage, executive functions, and reading time and preference in 4th to 6th graders. *The Korean Journal of Developmental Psychology*, 34(4), 109-131. <http://doi.org/10.35574/KJDP.2021.12.34.4.109>
- Kim, Kapim (2008). The Effects of Bibliotherapy on Elementary School Students' Self-Esteem and School Life Adjustment, Master's thesis, Aju University.
- Kim, Nayoung & Hwang, Hyeyoung (2020). Exploring variables on middle school students' academic

- engagement using Lasso Regression. *Journal of Education & Culture*, 26(6), 337-355.
<http://doi.org/10.24159/joec.2020.26.6.337>
- Kim, Taeyong & Park, Sunju (2014). A study on effects of elementary school students' use of smart phone upon reading and self-regulated reading. *Journal of The Korean Association of Information Education*, 18(3), 433-442. <http://doi.org/10.14352/jkaie.2014.18.3.433>
- Kim, Young-sik & Kim, Hoonho (2019). An inquiry for the predictive variables on the demand for the private tutoring utilizing machine learning approaches. *The Journal of Economics and Finance of Education*, 28(3), 29-52. <http://doi.org/10.46967/jefe.2019.28.3.29>
- Kim, Young-sik, Kim, Minseok, & Lee, Giljae (2019). Analysis on the factors predicting reading activities of high school students. *The Journal of Economics and Finance of Education*, 28(4), 137-156. <http://doi.org/10.46967/jefe.2019.28.4.137>
- Kwak, Kil-Yeo (2011). The Effect of Teacher's Reading Activity on Children's Self-Esteem and Reading Attitude in the Regional Child-Center. Master's thesis, National University in Suncheon.
- Lee, ByeongKi (2015). A study on the measurement tool development of reading motivation using self-determination theory. *Journal of the Korean Society for Library and Information Science*, 49(3), 311-332. <http://doi.org/10.4275/KSLIS.2015.49.3.311>
- Lee, Eun Ju, Kim, Min Jung, & Song, YeonJoo (2024). The effect of reading time and private education iime on academic engagement of middle school students: Mediating effect of self-esteem. *The Journal of Educational Research*, 22(2), 147-165.
<http://doi.org/10.31352/JER.22.2.147>
- Lee, Guk-Hee (2022). Effects of productive reading habit and constructual of reading on reading performance. *Journal of the Korea Contents Association*, 22(5), 565-575.
<http://doi.org/10.5392/JKCA.2022.22.05.565>
- Lee, Guk-Hee, Kim, Eunmi, Choi, Incheol, & Han, Dahye (2020). Effects of the interactions of reading preferences and reading construal levels on reading continuity and amount of reading. *Korean Journal of Social Science*, 39(1), 5-37.
<http://doi.org/10.18284/jss.2020.04.39.1.5>
- Lee, Gyeong-nam (2016). A study on factor analysis influencing on the confusion of reading comprehension. *Journal of Reading Research*, 40, 119-147.
<http://doi.org/10.17095/JRR.2016.40.4>
- Lee, Jaesun & Choi, Seoncheol (2020). Mediating effects of Grit on the relationship between smartphone usage time and reading time. *The Journal of Learner-Centered Curriculum*

- and Instruction, 20(7), 1009-1035. <http://doi.org/10.22251/jlcci.2020.20.7.1009>
- Lee, Young-Lim (2007). The Effects of Bibliotherapy Program on Self-esteem of the Adolescents. Master's thesis, Baekseok University.
- Ministry of Culture, Sports and Tourism (2023). 2023 National Reading Survey.
- Nam, Soojung (2023). The Mediating Effect of Academic Self-Efficacy in the Relationship between High School Students' Reading Activities, Parent-Child Communication, and Academic Achievement. Master's thesis, Pukyong National University.
- Park, Chan-heung (2022). Effect of longitudinal changes in reading time in elementary school students on career maturity and self-concept in the first year of middle school. *Journal of Elementary Korean Education*, 74, 123-142.
- Park, Hye-Sook (2008). A longitudinal study of the correlates among reading attitudes, reading behaviors, time spent on reading, and Korean reading achievements of middle school students. *The Journal of Educational Studies*, 39(3), 27-49.
- Park, Juhyeon, Lee, Myounggyu, & Park, Kyunghye (2015). The effect of reading aloud on the reading attitude and the self-esteem of elementary school students. *Journal of Korean Library and Information Science Society*, 46(2), 1-27.
<http://doi.org/10.16981/kliss.46.2.201506.1>
- Park, So Hyun (2020). A Study on the Predictors of Leisure Time of Elementary and Middle School Students. Master's thesis, Seoul National University of Education.
- Park, Soyoung & Chung, Hyewon (2019). Exploring factors affecting job satisfaction and organizational commitment of employees with machine learning technique. *Korean Journal of Educational Research*, 57(4), 183-213. <http://doi.org/10.30916/KERA.57.4.183>
- Park, Soyoung & Chung, Hyewon (2021). Exploring predictors affecting career development competence of high school students using random forest. *The Journal of Yeolin Education*, 29(1), 239-265. <http://doi.org/10.18230/tjye.2021.29.1.239>
- Park, Sung Jae & Shin, Insoo (2024). Exploring factors influencing youth reading in the era of Generative AI. *Journal of the Korean Society for Information Management*, 41(3), 171-203. <http://doi.org/10.3743/KOSIM.2024.41.3.171>
- Park, Sung Jae (2022). The impact of reading on self-efficacy and multicultural acceptance: Using multi-variate latent growth model. *Journal of Korean Library and Information Science Society*, 53(4), 293-318. <http://doi.org/10.16981/kliss.53.4.202212.293>
- Seo, Na La (2006). The Effects of a Bibliotherapy Program on the Self-Esteem and School Adjustment of Lower Graders at Elementary School. Master's thesis, Ewha Womans

University.

Seo, Sun-young (2007). The Effect of Reading-Guidance Program on Self-Esteem and Interpersonal Cognitive Problem Solving of Children. Master's thesis, Chinju National University.

Wang, Hyosung (2014). Exploring factors affecting adolescent readers' reading attitude: Based on reasons adolescent readers explained why they like or dislike reading. *Journal of Reading Research*, 33, 171-198. <http://doi.org/10.17095/JRR.2014.33.5>

