

## 머신러닝을 활용한 취학 전 아동의 내재화 및 외현화 문제행동에 미치는 예측 모형 및 영향 요인 탐색

김지윤<sup>1)</sup> 안새별<sup>2)</sup> 홍세희<sup>3)</sup>

### 요약

본 연구는 머신러닝 기법을 활용하여 취학 전 아동의 내재화 및 외현화 문제행동에 영향을 미치는 요인을 탐색하고 이들의 영향력을 살펴보고자 하였다. 한국아동패널(PSKC)의 7차 연도 자료 중 결측치를 제외한 총 2,150명의 아동과 데이터 전처리 과정을 통해 선정된 총 140개의 설명변수를 분석에 사용하였다. 분석 결과, 상위 3개 알고리즘을 기반으로, 내재화 문제행동의 경우 학습준비도와 친가의 사회적 지지가, 외현화 문제행동의 경우 어머니의 성역할 가치와 친가 및 이웃의 사회적 지지가 본 연구를 통해 취학 전 아동의 문제행동을 예측하는 주요 변수로 새롭게 도출되었다. 본 연구결과를 통해 문제행동의 관련 요인을 밝히고 문제행동 예방을 위한 유용한 정책적·교육적 시사점 및 전략을 구축할 수 있을 것으로 기대된다.

**주제어:** 문제행동, 내재화 문제행동, 외현화 문제행동, 취학 전 아동, 머신러닝

## I. 서론

유아기는 다른 시기에 비해 상대적으로 짧은 시간 안에 급격한 신체적, 정서적 성장을 하는 시기이며, 이때 이루어지는 발달과업 성취는 이후 시기까지 장기적인 영향을 미치므로 해당 시기에 바람직한 발달이 이뤄지는 것은 매우 중요하다(서석원, 2022). 하지만 공격성, 불안, 부적응 등의 문제행동은 유아의 성장에 필요한 경험 획득을 방해하여 긍정적인 방향의 발달 진행에 어려움을 겪게 만들 수 있고(김태련·강우선·김경은 외, 2003; 서석원, 2022), 학령기, 청소년기를 지나 성인기까지 그 영향이 지속해서 미쳐 이차적인 문제

- 1) 고려대학교 교육학과 교육측정·통계전공 박사수료
- 2) 고려대학교 교육학과 교육측정·통계전공 석사
- 3) 고려대학교 교육학과 교수

를 초래할 수 있다(신소정·이재모, 2011). 특히 만 6세의 유아는 초등학교 입학 앞둔 상태로, 해당 시기는 상급 학교로의 진학을 위한 학업적 준비와 더불어 낯선 환경에 적응하기에 필요한 사회성과 심리·정서적인 준비가 요구된다. 이러한 과정에서 안정적인 내적·외적인 상태 및 행동을 형성하는 데 실패한 아동은 초등학교 입학 이후 새로운 환경에 적응하는 데 어려움을 겪거나 학업에 따라가지 못하는 등 이후의 학교생활에서 부정적인 모습을 보일 우려가 큰 것으로 나타났다(이지영·이상희, 2019). 최근 문제행동을 보이는 유아기 아동이 증가하고 있으며(권수진·오새니·이상희, 2016), 문제행동이 발견되는 초기 연령이 낮아지는 추세(김영심·전정민·김정미, 2020)임을 고려해 보았을 때, 유아기 아동의 긍정적인 학습과 발달을 저해할 수 있는 문제행동을 정확하게 이해하여 유아의 바람직한 발달과 향후 학교적응을 도울 필요가 있다.

문제행동은 초기에 사회 부적응으로 불리며 갈등 상황 시 욕구 조절에 실패하여 나타나는 부적절한 반응으로 이해되었다(조정숙, 2015). 이후 해당 용어와 함께 역기능적 행동, 행동 문제 또는 문제행동이 통용되어 쓰였고, 정서와 행동 간의 관련성이 인식됨에 따라 정서 및 행동장애로 불리기도 하였다(이성봉·방명애·김은경 외, 2014). 이처럼 통용되는 명칭은 여러 개이나, 문제행동은 기대되는 행동에서 벗어난 부적절한 정서적 및 행동적 반응으로 정의될 수 있다(Garber, 1984). 특히 문제행동의 주체가 유아기 아동일 때 이들의 연령과 발달 수준의 특수성을 고려하여 이해될 필요가 있다. 유아의 문제행동은 아동의 발달과 놀이를 방해하고 타인에게 피해를 주거나 사회 규범에서 벗어난 행동으로 정의될 수 있고(Kaiser & Rasminsky, 2007), 해당 연령에 맞는 발달 수준에 부적합한 행동, 새로운 환경에 정상적으로 반응하지 못해 이루어진 부적응적 행동, 또는 보호자의 지도를 벗어난 행동을 말한다(김연·한태숙·황혜정, 2004).

문제행동은 불안, 위축 등 개인이 내적으로 경험하는 정서 및 심리적 문제인 내재화 문제행동과 공격성, 과잉행동 등 겉으로 드러나는 행동적 문제인 외현화 문제행동으로 구분될 수 있다. 내재화 문제행동의 경우 문제가 되는 상태가 눈으로 쉽게 보이지 않는다는 점 때문에 초기에 발견되기 어렵고(정미라·이정은·박혜성, 2013), 적당한 시기 내 해결되지 않는다면 아동의 일과를 방해하여 성공적인 성취를 해친다고 알려져 있다(Poulou, 2015). 반면 외현화 문제행동은 해당 양상이 외적으로 드러나기에 조기 발견이 쉬우나(정미라·이정은·박혜성, 2013), 적절하게 완화되지 않을 시 바람직한 관계 형성과 환경 적응에 실패할 개인성이 높아지는 것으로 나타났다(Greenberg, Speltz & Deklyen, 1993; 김영미·송하나, 2018에서 재인용). 따라서 유아의 바람직한 성장을 위해 아동의 내재화 및 외현화 문제행동 각각에 대한 영향요인을 파악할 필요가 있으며, 이를 기반으로 유아기 문제행동

을 예방하고 완화하기 위한 접근을 도출해 낼 수 있을 것이다.

유아의 문제행동 영향요인을 규명한 기존의 연구를 살펴보면, 아동 개인의 내적 요인뿐 아니라 이들을 둘러싼 다양한 변인에 따라 문제행동이 달라질 수 있음을 알 수 있다. 먼저 개인의 내적 요인을 살펴보면, 아동의 성별에 따라 내재화·외현화 문제행동 수준이 다르게 나타났고(원영미·박혜원·이귀옥, 2004; 임현주, 2015), 윤갑로(1994)에 의하면 아동의 출생 순위와 관련하여 둘째보다는 첫째가 문제행동을 더 많이 보이기도 하였다(이은주·이기연, 1995에서 재인용). 유아의 일상적 스트레스 역시 이들의 내재화 및 외현화 문제행동에 영향을 미치는 것으로 나타나, 스트레스 지수를 낮춤으로써 문제행동을 감소시키는 전략이 필요하다고 보았다(최선녀·최항준, 2017). 반면, 유아가 행복할수록 이들의 문제행동이 보고되는 빈도가 낮아(김지혜, 2011), 긍정적 정서는 문제행동을 낮추는 요인으로 추측된다. 특히 다양한 선행연구에서 사회적 유능감은 유아의 내재화 문제행동을 낮추는 요인으로 나타났다(서석원, 2022; 우수경, 2017; 이찬숙·현은자, 2008). 사회적 유능감은 자신의 정서를 적절히 통제하고 타인과의 관계를 이해하는 능력으로, 자신의 감정을 잘 조절하고 사회적 관계를 잘 형성함으로써 내재화 문제행동의 발현이 최소화할 수 있다고 보고되었다(서석원, 2022). 따라서 내재화 문제행동은 유아가 내적으로 겪게 되는 문제 양상이라는 점에서 유아의 정서 및 심리 상태와 관련한 변인이 내재화 문제행동의 주요 예측 요인으로 도출될 수 있다고 추측해 볼 수 있다.

특히 학습준비도는 문제행동 중에서도 외현화 문제행동을 낮추는 주요 요인으로 제시되었다(하민경, 2022). 해당 문제행동은 유아가 타인과 관계를 형성할 때 문제가 되는 행동 양상이라는 점에서 학업 수행 능력뿐 아니라 의사소통 및 인지능력까지 포함하는 다차원적 개념인 학습준비도(정진나, 2020)와 깊은 연관이 있음을 추측해 볼 수 있다. 산만한 학습태도는 주의집중을 방해하고 낮은 사회정서 능력은 공격성 수준을 높이는 것으로 나타나, 언어 및 인지, 사회정서 능력의 적절한 발달을 촉진함으로써 외현화 문제행동을 예방할 수 있다(하민경, 2022). 이와 유사하게 인지능력과 의사소통 수준이 높은 아동일수록 집중력이 좋고 타인과 더욱 안정적인 상호작용이 가능하다는 점에서 낮은 수준의 외현화 문제행동을 보이는 것으로 나타났다(이진숙·한지현·박혜원, 2016). 따라서 외현화 문제행동은 사회관계 속에서 중요하게 작용할 수 있는 인지 및 언어 능력, 타인과의 상호작용과 관련된 변인이 주요 예측요인으로 도출될 수 있을 것이다.

유아의 문제행동은 부모 및 교사와 관련된 요인과의 매우 밀접한 관련이 있었다. 특히 부모의 양육행동은 다수의 선행연구에서 유아 문제행동의 주요 예측요인으로 제시되었다. 부모의 부정적 양육태도는 유아의 문제행동을 높이는 데 영향을 주었고(김효정·조동혁·조희

준, 2023). 의존적 또는 거부적 양육행동은 유아 문제행동의 원인으로 나타났다(이찬숙·현은자, 2008). 또한, 부모가 함께 자녀를 양육하고자 할수록 긍정적인 부모-자녀 간 상호작용이 이뤄지고 안정적인 가정환경이 구축됨으로써 유아의 외현화 문제행동을 낮추는 데 도움이 되었다(유다정·한세영, 2022). 특히 부모의 정서 및 심리 상태는 유아의 문제행동과 깊은 연관이 있다고 연구되었다. 어머니의 스트레스는 자녀의 문제행동에 부정적인 영향을 주는 것으로 나타났다(서석원, 2019; 임현주, 2015), 아버지의 스트레스 또한 유아의 문제행동을 높이는 요인으로 작용하는 것으로 확인되어(이지예·신경순·김영희, 2017), 부모의 부정적인 심리 상태는 자녀의 문제행동에 의미 있는 영향을 지님을 추측해 볼 수 있다.

유아기 아동은 적지 않은 시간 동안 육아 기관 교사의 지도를 받으며 성장하기 때문에 이들로부터 받는 영향을 무시할 수 없다. 교사와 관련된 특징을 아동의 문제행동과 연결한 선행연구를 살펴보면, 교사와 유아 사이의 친밀한 관계는 문제행동을 낮추는 데 도움을 주었고(신유림, 2021), 교수효능감이 높은 교사로부터 지도받은 유아는 문제행동을 덜 보이는 것으로 보고되었다(염수진·이희영·최태진, 2022). 이외에도 가정환경과 보육환경의 특성 또한 유아의 문제행동에 의미 있는 영향을 미치는 요인으로 제시되었다(김수정·정익중, 2015; 하영례·정효은, 2009).

문제행동을 다룬 기존 선행연구는 대부분 내재화 및 외현화 문제행동을 하나로 묶어 해당 영향요인을 알아보는 데 초점을 맞추었다(김효정·조동혁·조희준, 2023; 임현주, 2015; 최선녀·최향준, 2017). 다만, 기존 연구에 의하면 내재화 및 외현화 문제행동은 서로 다른 특징을 보이는 것으로 알려져 있다(정미라·이정은·박혜성, 2013; Poulou, 2015). 따라서 문제행동은 하나의 부적응 행동으로 다뤄지기보다는 내재화 또는 외현화 문제행동으로 따로 구별되어 다뤄질 필요가 있다. 또한, 내재화 및 외현화 문제행동 중 하나만을 선택하여 분석하는 경우도 많았으나(서석원, 2022; 정진나, 2020; 하민경, 2022), 두 문제행동을 동시에 분석하여 영향요인을 비교한 선행연구는 드물었다. 이에 본 연구는 두 문제행동을 구분하여 각각의 예측요인을 동시에 도출해냄으로써 둘의 공통점과 차이점을 밝혀내고자 한다. 또한, 문제행동의 예측요인을 밝힌 선행연구는 대부분 개인, 부모, 교사요인과의 관련성에 집중하였을 뿐, 패널자료에서 활용 가능한 가정 및 육아기관 등의 환경적 요인을 투입한 연구는 비교적 많지 않았다. 유아기 아동은 부모와 교사와 같은 주변 인물뿐만 아니라 이들이 속한 가정과 육아기관 또한 의미 있는 영향을 받을 수 있으므로 해당 변인과의 관계성을 확인해 볼 필요가 있다. 따라서 본 연구는 개인, 부모, 교사요인과 더불어 가정환경과 육아기관 관련 변수와의 관계를 확인해 봄으로써 해당 환경 요인이 유아기 아동의 내재화 및 외현화 문제행동의 주요한 예측요인이 될 수 있는지를 확인하고자 한다.

이에 본 연구는 취학 전 아동의 내재화 및 외현화 문제행동 각각의 주요 예측요인을 탐색하기 위하여 머신러닝 기법을 적용하였다. 머신러닝은 많은 수의 변수를 동시에 투입함으로써 핵심 변수와의 관계와 예측요인의 중요도를 확인할 수 있다는 강점을 지닌다. 따라서 본 연구는 내재화 및 외현화 문제행동별로 예측에 우수한 성능을 보이는 상위 3개 알고리즘을 각각 선정하여 해당 알고리즘에서 문제행동 예측에 중요도를 기반으로 기존의 연구에서 다루어지지 않은 새로운 변수를 뽑아냄으로써 취학 전 아동의 문제행동에 관한 유용한 정보를 도출해내고자 하였다. 또한, 주요 예측요인을 비교함으로써 내재화 및 외현화 문제행동에 영향을 미칠 수 있는 공통적인 요인과 각 문제행동에 영향을 주는 요인을 확인하여, 문제행동 예방을 위한 접근과 전략에 대해 제안하고자 한다. 본 연구는 머신러닝 기법을 통해 유아의 내재화 및 외현화 문제행동에 영향을 주는 새로운 주요 변수를 도출함으로써 아동의 문제행동에 관한 유용한 가이드라인을 제시해 줄 수 있을 뿐만 아니라, 관련 후속 연구에 새로운 방향성이나 다각적인 시사점을 제공할 수 있으리라 기대한다. 더불어, 문제행동 관련 선행연구에서는 기존에 밝혀진 예측변인을 사용함으로써 문제행동 예방을 위한 전략도 해당 변인에만 한정되어 제안에 한계가 있었다. 반면 본 연구는 기존 연구에서 밝혀지지 않은 문제행동 예측변인을 추가로 도출해냄으로써 유아기 아동의 문제행동에 도움이 되는 정책 또한 새롭게 제안할 수 있을 것이다. 본 연구의 연구 문제는 다음과 같다.

연구 문제 1. 유아기 아동의 내재화 문제행동 예측에 영향을 주는 중요 변수는 무엇인가?

연구 문제 2. 유아기 아동의 외현화 문제행동 예측에 영향을 주는 중요 변수는 무엇인가?

## II. 연구방법

### 1. 연구대상

본 연구는 학령 전 아동의 내재화 및 외현화 문제행동 예측에 미치는 영향변수를 탐색하기 위해 한국아동패널(Panel Study on Korean Children; PSKC)의 7차 연도(2014년) 자료를 사용하였다. 한국아동패널은 육아정책연구소에서 2008년에 출생한 신생아를 대상으로 2027년까지 매해 아동, 부모, 가족, 학교, 지역사회 등에 대해 추적 조사가 실시되고 있는 패널연구이며, 층화다단계 표집을 실시함으로써 단순임의추출보다 신뢰성이 높은 표본을 선정하였다고 평가되고 있다. 이를 통해 각 시기의 아동에 대한 성장 및 발달 특성, 부모의 양육 특성, 육아 및 교육기관의 기능과 효과의 영향, 그리고 이들의 관계 탐색에 유

용하게 활용할 수 있다. 본 연구는 학교급 변화가 발생하기 전인 7차 연도 자료를 선택하여, 이 시기에 형성된 다양한 요인이 아동의 내재화 및 외현화 문제행동에 미치는 영향력을 확인하고자 하였다. 머신러닝을 활용한 선행연구와 동일한 방식을 적용하기 위하여(정혜원·김예림·박소영, 2020; 송은정·김지윤·안새별 외, 2022), 본 연구의 분석에 투입되는 모든 변수에서 응답 자료의 결측률이 50% 이상인 표본을 제외하였다. 이에 따라 7차 연도에 포함된 총 2,150명의 아동 중 남자는 50.7%, 여자는 49.3%이었다.

## 2. 측정도구

### 가. 결과변수: 문제행동

본 연구에서 종속변수로 사용된 취학 전 아동의 문제행동은 아동의 주양육자가 자녀의 사회 적응 및 정서 행동에 관한 문제행동을 평가하는 척도로, 그중 만 1.5-5세 영유아를 대상으로 하는 유아 행동 평가척도(Child Behavior Checklist, CBCL for Ages 1.5-5)가 사용되었다. 이는 Achenbach와 Rescorla(2000)가 개발한 아동·청소년의 문제행동을 부모나 교사의 보고를 통해 평가할 수 있는 척도 ASEBA(Achenbach System of Empirically Based Assessment) 중 유아용 평가척도를 우리나라의 정서에 맞게 번역 및 표준화한 척도이다(김영아·이진·문수중 외, 2009에서 재인용). 해당 척도를 통해 신체적, 정신적 발달이 급격히 이루어지고 발달의 기초를 형성하는 중요한 시기에 아동의 정상적 발달을 지연시키거나 어렵게 하는 문제를 행동적 특성으로 파악함으로써, 조기에 필요한 개입 및 중재를 판단할 수 있다. 유아들에게 관찰되는 정서·행동 문제에 관한 총 100개의 문항을 3점 척도(0: 전혀 아니다, 1: 가끔 그렇다, 2: 자주 그렇다)로 평가하며, 점수가 낮을수록 문제의 정도가 약함을 의미한다. 문제행동 증후군은 총 8개의 영역(정서적 반응성, 불안/우울, 신체증상, 위축, 주의집중문제, 공격행동, 수면문제, 기타 문제)으로 구성되며, 해당 영역들의 상위범주로 내재화 척도와 외현화 척도로 나뉜다. 내재화 척도에는 정서적 반응성, 불안/우울, 신체증상, 위축 영역이 포함되고, 외현화 척도에는 주의집중문제, 공격행동 영역이 포함된다. 본 연구에서 사용된 문제행동의 Cronbach's alpha 신뢰도는 내재화 척도가 .873, 외현화 척도가 .881로 나타났다.

### 나. 설명변수

취학 전 아동의 내재화 및 외현화 문제행동 예측요인 탐색을 위한 설명변수로 한국아동 패널의 7차 연도에 조사된 변수를 사용하였다. 이 중 개인 식별 변수, 횡단 및 종단 가중치

와 같이 분석에 불필요하거나 코드형 및 주관식 응답 변수와 같이 분석에 포함하기 어려운 변수를 제외하였다. 따라서 아동, 부모, 가구, 교사, 육아 지원기관의 특성 변인에 관한 총 142개의 변수를 분석의 대상으로 하였다. 분석에 포함된 변수 중 일부 문항은 결과 해석의 용이함을 위하여 역코딩 처리하였다. 범주형 변수를 제외한 모든 변수의 경우, 각 변수에 포함된 문항의 평균값을 산출하여 분석에 활용하였다. 분석에 투입될 변수 중 응답률의 결측치가 50% 이상을 포함하여 결측치 대체 방법 적용이 어려운 2개의 변수를 삭제하고, 나머지 변수를 대상으로 라이트 그래디언트 부스팅(Light Gradient Boosting Machine) 모형의 MICE(Multiple Imputation using Chained Equation) 방식을 10회 반복하여 결측치를 대체하였다. 변수 간 상관관계가 .900 이상인 경우 해당 변수를 제거하여 다중공선성 문제를 방지하였으며, 정규화 방법을 통해 모든 변수가 동일한 척도선상에 있도록 재척도화하였다. 이러한 데이터 전처리 과정을 통하여 최종적으로 총 140개의 변수가 취학 전 아동의 내재화 문제행동과 외현화 문제행동 예측의 설명변수로 각각 투입되어 사용되었다. 본 연구에서 최종 사용된 설명변수는 다음의 <표 1>과 같다.

<표 1> 설명변수의 영역별 분류

변수	변수 내용	
인구학적 특성	성별, 월령, 출생순위	
일상생활 특성	숙면 여부, 수면 시간, 전자기기 놀이일, 미디어 이용 정도, 실내·실외 신체활동, 일상적 스트레스	
아동 특성	신체 및 건강 특성	사고/중독으로 인한 치료 여부, 입원 경험, 장애 판정 여부, 간접 흡연 여부
	사회 및 정서 발달	성장·발달·적응에 관한 전반적 발달(부/모/교사), 전반적 행복도, 육아기관 선호도, 교사 선호도, 자아존중감, 주관적 행복, 언어·수리 영역의 학업적 능력(모/교사), 학습준비도(모/교사), 사회적 유능감(모/교사), 또래놀이 상호작용
부모 특성	개인 및 부부 특성	연령, 최종 학력, 우울, 자아존중감, 주관적 행복감, 일상적 스트레스, 배우자의 일상적 스트레스, 삶의 만족도, 결혼 만족도, 부부 갈등, 성역할 가치, 취업·학업 상태, 일자리에 대한 만족도
	신체 및 건강 특성	음주 여부 및 횟수, 음주량, 과음 횟수, 흡연 여부, 장애 유무
	양육 특성	양육행동(온정적/통제적), 양육효능감, 양육스트레스, 공동양육, 양육 부담, 부모-자녀 상호작용, 가족 상호작용, 사회적지지(친가/외가/친구 및 동료/이웃), 육아기관 방문 횟수, 육아기관 행사 참여 여부, 육아기관 선호도, 육아기관 교사 선호도, 육아기관을 보내며 받는 지원 만족도, 기관 생활 관심 정도

변수		변수 내용
가구 특성	가정환경 특성	가정환경의 질(학습자료/언어자극/물리적 환경/반응성/학습자극/모방학습/다양성/수용성)
	일반적 특성	거주 지역 규모, 거주 지역 6개 권역 구분
교사 특성	개인 특성	성별, 연령, 최종학력, 전공, 자격내용
	교수 특성	업무 스트레스, 교실환경, 교수 효능감, 교사-아동 상호작용, 전반적 행복감
육아기관 특성		설립 유형, 학급 아동 수 적정성, 교사 대 아동 비율, 교실 규모 적절성, 실외 놀이 공간 유무, 실내 대근육 활동 공간 유무

### 3. 분석 방법

본 연구는 학령 전 아동의 내재화·외현화 문제행동 예측에 영향을 미치는 주요 변인을 도출하기 위해 머신러닝 분석 방법을 수행하였다. 머신러닝 기법은 대량의 자료로부터 특정 규칙 및 패턴을 학습하여 이를 새로운 자료 세트에 대한 적절한 작업으로 수행하는 일련의 알고리즘 또는 프로세스를 의미하며(Dangeti, 2017), 머신러닝 기법을 통해 다수의 설명변수를 동시에 투입하여 주요 핵심 변수 선정이 가능하다(유진은·김형관·노민정, 2020). 기존의 통계적 방법을 적용하여 하나의 모형에 수백 개의 변수를 포함하는 경우, 비수렴 및 과적합 등의 문제가 발생할 수 있다(유진은·김형관·노민정, 2020). 따라서 본 연구에서는 학습 및 평가 데이터에 대한 분리, 교차 타당화, 추정된 계수의 정규화 등 다양한 접근법을 기반으로 이러한 문제를 최소화할 수 있는 머신러닝 기법을 적용하였다. 특히 본 연구에서는 머신러닝 모델 개발 작업을 자동화하여 생산성과 효율을 높여주는 프로세스인 자동 기계학습(Automated Machine Learning, AutoML)(문용혁·신익희·이용주 외, 2019)을 수행할 수 있는 파이썬(Python) 라이브러리 중 하나인 PyCaret을 활용하였다. PyCaret은 머신러닝 모델 개발의 각 단계인 데이터 전처리, 모델 선택, 하이퍼 파라미터(Hyper-parameter) 최적화 등의 작업을 자동화해 주는 라이브러리로, 주어진 학습 데이터에 대하여 분류(classification), 회귀(regression), 군집(clustering), 이상치 탐지(anomaly detection) 등 다양한 머신러닝 모델의 알고리즘을 지원한다(Ali, 2020; Gain & Hotti, 2021). 본 연구에서는 PyCaret이 제공하는 20여 개 회귀분석 모델들의 성능을 비교하였다.

본 연구의 분석 과정은 다음과 같다. 첫째, 회귀분석 모델을 적용하기 전 데이터 전처리 과정을 거친 연구자료를 8:2 비율로 임의 분할하여 사용함으로써(학습 데이터: 1,720명, 평가 데이터: 430명), 학습 데이터와 평가 데이터가 독립적이지만 동시에 동일한 확률분포

를 다르게 하였다(Dangeti, 2017). 둘째, 교차검증(Cross-Validation)을 10회 반복하는 10겹 교차검증(10-fold CV)을 바탕으로 최적의 하이퍼 파라미터 탐색을 위해 그리드검색(GridSearch)을 실시하였다. 그리드검색은 연구자가 설정한 값 중 가장 최적화된 하이퍼 파라미터 조합을 찾아주는 방법이며, 이를 통해 모델에서 제공하는 최적의 하이퍼 파라미터값을 도출하여 더욱 정확한 학습을 기반으로 모델의 성능을 향상할 수 있다(박세준·변영철, 2021). 셋째, 이러한 과정을 통해 도출된 최적의 하이퍼 파라미터 조합으로 추정된 예측 모형을 다양한 평가 지수를 기반으로 성능을 비교하여, 우수한 성능이 확인된 상위 3개 알고리즘을 선정하였다. 우수한 성능을 보이는 단 하나의 알고리즘을 선정하고, 해당 알고리즘에서 주요 변수를 도출하는 방식으로 진행한 기존 머신러닝 연구와는 달리(고은경, 2020; 박소영·정혜원, 2020), 본 연구에서는 상위 3개 알고리즘에서 도출한 결과를 통합·제시함으로써 보다 타당하고 객관적으로 주요 변수를 탐색하고자 하였다. 이후 알고리즘별로 최적화된 알고리즘에 대해 10겹 교차검증과 그리드검색을 다시 한번 적용하여 최종적으로 알고리즘의 성능을 평가하였다. 마지막으로, 선정된 상위 3개의 알고리즘에서 산출되는 회귀 계수/중요도 지수를 기준으로, 취학 전 아동의 문제행동 예측에 영향을 주는 설명 변수를 탐색하였다. 여러 알고리즘에서 공통으로 탐색되는 주요 변수 파악을 위하여, 알고리즘별로 설명변수의 회귀 계수/중요도 지수를 기반으로 중요도 순위를 매긴 후 그 순위에 대한 평균값을 산출하여 문제행동에 중요한 영향을 미치는 설명변수를 도출하였다. 본 연구는 데이터 전처리와 머신러닝 분석을 위해 프로그래밍 언어 Python 3.10의 PyCaret 2.3.10 패키지를 사용하였다.

### III. 연구결과

#### 1. 예측모형 성능 평가

취학 전 아동의 내재화 및 외현화 문제행동 예측을 위하여 140개의 설명변수를 투입하고 20여 개의 회귀 모델 중 상위 5개의 성능을 평가한 결과는 <표 2>와 같다. 특히 모든 알고리즘에서 전반적으로 낮은  $R^2$ 값이 확인되었는데, 해당  $R^2$ 의 값은 분석에 사용된 140개의 설명변수 모두를 투입한 모형의 예측력이다. 하지만 모형에 투입되는 설명변수의 수가 증가할수록 회귀모형의 복잡도가 높아져 모형의 예측 성능이 저하될 가능성이 커진다(류나현·김형석·강필성, 2016). 이러한 이유로 본 연구에서는  $R^2$ 를 모형의 예측력이나 설

명력이 아닌 회귀모형 알고리즘 간 상대적인 성능 비교를 위한 평가 지수로 활용하였다. 학습 데이터(1,720명)를 활용하여 적합한 수준의 성능이 확인된 상위 5개 알고리즘을 확인한 결과, 내재화 문제행동의 경우 RMSE를 기준으로 Bayesian Ridge의 예측 오차가 5.265로 가장 작은 것으로 확인되었다. Gradient boosting Regressor 5.324, Extra Trees Regressor 5.373, Elastic Net 5.392, Random Forest Regressor 5.408의 순으로 작은 예측 오차를 보였다. 다음으로 외현화 문제행동의 경우, RMSE를 기준으로 내재화 문제행동과 동일하게 Bayesian Ridge의 예측 오차가 4.651로 가장 작은 것으로 나타났다. 다음으로 Gradient boosting Regressor 4.708, Random Forest Regressor 4.719, Extra Trees Regressor 4.739, Light Gradient Boosting Machine Regressor 4.763의 순으로 예측 오차가 작은 것으로 나타났다. 또한  $R^2$ 를 기준으로 살펴보았을 때도, 내재화 문제행동과 외현화 문제행동 모두 RMSE를 기준으로 살펴본 경우와 동일한 순서로 예측력이 높게 나타났다.

이러한 결과를 기반으로 본 연구에서는 취학 전 아동의 문제행동을 예측한 회귀분석 알고리즘 중 예측 성과가 가장 우수한 상위 3개 모델을 내재화 문제행동과 외현화 문제행동 별로 각각 선정하였다.

〈표 2〉 회귀 모델 간 성능 평가 결과

모델	MAE	MSE	RMSE	$R^2$	RMSLE	MAPE
<b>내재화 문제행동</b>						
Bayesian Ridge	4.116	27.943	5.265	.182	.839	1.028
Gradient Boosting Regressor	4.179	28.462	5.324	.158	.843	1.058
Extra Trees Regressor	4.234	29.051	5.373	.147	.862	1.077
Elastic Net	4.223	29.323	5.392	.144	.867	1.069
Random Forest Regressor	4.277	29.436	5.408	.134	.869	1.110
<b>외현화 문제행동</b>						
Bayesian Ridge	3.575	21.762	4.651	.198	.832	.983
Gradient Boosting Regressor	3.627	22.286	4.708	.174	.838	.973
Random Forest Regressor	3.674	22.412	4.719	.173	.859	1.008
Extra Trees Regressor	3.688	22.555	4.739	.166	.855	1.022
Light Gradient Boosting Machine	3.698	22.823	4.763	.157	.861	.999

취학 전 아동의 내재화 및 외현화 문제행동에 대한 예측력을 높이기 위하여, 우수한 성능을 보인 상위 3개의 알고리즘을 대상으로, 각 알고리즘에 그리드검색과 10겹 교차검

증을 다시 한번 적용하여 재 최적화하고 알고리즘의 성능을 최종적으로 비교하였다. 이러한 결과는 <표 3>에 제시하였으며, 10겹 교차검증을 통해 산출된 평가 지수의 평균값을 제시하였다. 우선, 내재화 문제행동의 예측력에서 우수한 성능을 보고한 Bayesian Ridge, Gradient Boosting Regressor, Extra Trees Regressor의 경우 재최적화하기 전과 후의 성능이 모든 평가 지수 값 및 순위에서 크게 차이가 없었다. 하지만 외현화 문제행동의 경우, RMSE와  $R^2$  기준으로 하였을 때 Bayesian Ridge, Random Forest Regressor, Gradient Boosting Regressor 순으로 예측 오차가 작고 예측력이 높은 것으로 확인이 되어, 이전에 도출된 알고리즘 간 성능 평가 순위와 다르게 나타났다. 이는 다시 한번 최적화하는 과정을 통해 Random Forest Regressor가 Gradient Boosting Regressor보다 더욱 우수한 예측 성능을 가지게 되고, 이에 따라 알고리즘의 성능 평가 순위가 변동되었음을 의미한다.

<표 3> 최적화 이후 회귀 모델 간 성능 평가 최종 결과

모델	MAE	MSE	RMSE	$R^2$	RMSLE	MAPE
<b>내재화 문제행동</b>						
Bayesian Ridge	4.116	27.942	5.265	.182	.839	1.028
Gradient Boosting Regressor	4.176	28.707	5.345	.153	.848	1.047
Extra Trees Regressor	4.196	29.016	5.367	.151	.862	1.056
<b>외현화 문제행동</b>						
Bayesian Ridge	3.574	21.761	4.651	.198	.832	.983
Gradient Boosting Regressor	3.687	22.427	4.725	.171	.862	1.011
Random Forest Regressor	3.638	22.299	4.708	.178	.849	.980

## 2. 설명변수의 중요도

재 최적화 과정을 통해 우수한 성능을 보고한 상위 3개 알고리즘을 기반으로, 투입된 140개의 설명변수 중 취학 전 아동의 내재화 및 외현화 문제행동을 예측하는 주요 요인을 탐색하기 위하여 설명변수의 중요도를 살펴보았다. 즉, 내재화 문제행동의 경우 Bayesian Ridge, Gradient Boosting Regressor, Extra Trees Regressor 알고리즘, 외현화 문제행동의 경우 Bayesian Ridge, Gradient Boosting Regressor, Random Forest Regressor 알고리즘 각각에서 회귀 계수 및 중요도 지수를 기반으로 설명변수 예측력의 정도에 대해 순위를 부여하였다. 이후 상위 3개 알고리즘에서 부여된 순위 값의 평균을 산

출하여 설명변수의 중요도를 확인하였다. 이러한 과정을 바탕으로 취학 전 아동의 문제행동을 예측하는 상위 10개의 중요 요인이 도출되었으며, 내재화 및 외현화 문제행동의 최종 선택된 설명변수는 <표 4>, <표 5>와 같다.

<표 4> 내재화 문제행동 - 설명변수의 중요도 순위 평균

설명변수	BR	GBR	ETR	평균 순위값	변수설명
학습준비도(모 평정)	2	1	2	1.667	22문항(4점 척도), 예시: 수업 활동에 열정적으로 흥미를 가지고 참여한다
사회적 유능감(모 평정)	3	2	1	2	32문항(3점 척도), 예시: 다른 사람에게 먼저 말을 건넨다
모의 양육 스트레스	4	3	3	3.333	11문항(5점 척도), 예시: 아이로부터 도망치고 싶을 때가 있다
출생순위	1	6	5	4	1문항(연속 척도)
모의 우울	7	4	4	5	6문항(5점 척도), 예시: 불안하십니까? 1문항(7점 척도), 예시: 어머니가 보시기에 OO(이)의 평소 모습(행복도)은 어떠합니까?
전반적 행복도(모 평정)	9	5	6	6.667	1문항(4점 척도), 전반적 발달 중 기관 적응에 대해 평정
전반적 발달:기관 적응(모 평정)	5	9	10	8	13문항(5점 척도), 예시: 외로울 때 솔직히 털어놓고 의지할 수 있다
친가의 사회적 지지	6	8	14	9.333	1문항(4점 척도)
일상적 스트레스(모 평정)	12	10	9	10.333	1문항(4점 척도)
모의 일상적 스트레스	14	20	12	15.333	1문항(4점 척도)

주: BR - Bayesian Ridge, GBR - Gradient Boosting Regressor, ETR - Extra Trees Regressor

<표 5> 외현화 문제행동 - 설명변수의 중요도 순위 평균

설명변수	BR	GBR	RFR	평균 순위값	변수설명
사회적 유능감(모 평정)	1	2	1	1.333	32문항(3점 척도), 예시: 다른 사람에게 먼저 말을 건넨다
모의 양육 스트레스	2	1	2	1.667	11문항(5점 척도), 예시: 아이로부터 도망치고 싶을 때가 있다
학습준비도(모 평정)	5	3	5	4.333	22문항(4점 척도), 예시: 수업 활동에 열정적으로 흥미를 가지고 참여한다
언어 영역 학업 능력(모 평정)	4	5	6	5	28문항(5점 척도), 예시: 다소 복잡한 문장 구조를 사용한다

설명변수	BR	GBR	RFR	평균 순위값	변수설명
친가의 사회적 지지	8	4	4	5.333	13문항(5점 척도), 예시: 외로울 때 솔직히 털어놓고 의지할 수 있다
이웃의 사회적 지지	15	10	8	11	13문항(5점 척도), 예시: 외로울 때 솔직히 털어놓고 의지할 수 있다
또래놀이 상호작용(교사 평정)	20	6	7	11	30문항(4점 척도), 예시: 친구를 돕는다
모의 우울	18	13	3	11.333	6문항(5점 척도), 예시: 불안하십니까?
모의 부모 간 공동양육	16	16	13	15	16문항(7점 척도), 예시: 배우자에게 아이에 대한 칭찬이나 긍정적인 말을 한다
모의 성역할 가치	9	19	18	15.333	6문항(4점 척도), 예시: 취업을 한 어머니도 전업 주부 어머니만큼 자녀와 친밀한 관계를 가질 수 있다

주: BR - Bayesian Ridge, GBR - Gradient Boosting Regressor, RFR - Random Forest Regressor

먼저, 내재화 문제행동의 경우, 아동의 학습준비도, 사회적 유능감, 출생 순위, 전반적 행복도, 기관 적응, 일상적 스트레스와 어머니의 양육 스트레스, 우울, 일상적 스트레스, 친가의 사회적 지지 변수들이 예측에 중요한 영향을 미치는 변수로 도출되었다. 또한 외현화 문제행동의 경우 아동의 사회적 유능감, 학습준비도, 언어 영역의 학업 능력, 또래놀이 상호작용과 어머니의 양육 스트레스, 우울, 부모의 공동양육, 성역할 가치, 친가 및 이웃의 사회적지지 변수들이 주요 변수로 나타났다. 이러한 분석 결과는 사회적 유능감, 학습준비도, 일상적 스트레스, 어머니의 양육 스트레스, 우울 등을 아동 문제행동의 주요 예측 변수로 살펴본 기존의 선행연구와 일치한다(연은모·윤해옥·최효식, 2016; 임정수·이진영·남보영, 2022; 최은정·김은향, 2019; 최지은, 2020). 반면, 이론적 고찰을 통해 살펴본 중요 변수와 비교해 본 결과, 학습준비도, 친가 및 이웃의 사회적지지, 어머니의 성역할 가치는 기존의 연구에서 아동의 내재화 및 외현화 문제행동과의 관계가 규명되지 않은 변수로 나타나, 본 연구에서 학령 전 아동의 내재화 및 외현화 문제행동을 예측하는 주요 변수로 새롭게 도출되었다.

#### IV. 논의 및 결론

본 연구는 한국아동패널 7차 자료를 활용하여 취학 전 아동의 내재화 및 외현화 문제행동 예측에 우수한 성능을 보이는 머신러닝의 회귀모형 알고리즘을 비교 및 선정하고, 선정

된 알고리즘을 최적화하여 아동의 문제행동을 예측하는 데 영향을 미치는 주요 변수를 탐색하고자 하였다. 또한 기존의 선행연구에서 도출되지 않았던 새로운 변인을 도출하고, 이러한 변수를 통해 취학 전 아동의 문제행동 교정에 도움을 주는 시사점을 제공하고자 하였다. 본 연구의 주요 결과와 논의는 다음과 같다.

첫째, 취학 전 아동의 문제행동 예측을 위한 다양한 머신러닝 회귀모형들의 성능 비교를 수행한 결과, RMSE를 기준으로 내재화 문제행동의 경우 Bayesian Ridge, Gradient Boosting Regressor, Extra Trees Regressor 알고리즘 순으로, 외현화 문제행동의 경우 Bayesian Ridge, Gradient Boosting Regressor, Random Forest Regressor 알고리즘 순으로 예측 오차가 작은 것으로 확인되었다. 예측력을 나타내는 결정계수  $R^2$  값을 기준으로 살펴보았을 때도 위와 동일한 순서로 예측력이 높았다. 이후, 우수한 성능을 보인 상위 3개의 알고리즘을 대상으로 그리드검색과 10겹 교차검증을 다시 한번 적용하여 알고리즘의 성능을 최종적으로 비교한 결과, 내재화 문제행동의 경우 재 최적화하기 전과 후의 예측 성능에 대한 상위 3개 알고리즘의 평가 순위에서 차이가 없었다. 반면, 외현화 문제행동의 경우, 재 최적화 과정을 통해 Random Forest Regressor의 예측 성능이 향상되었으며, 그 결과 Gradient Boosting Regressor보다 우수한 성능을 보였다. 재최적화 과정으로 인해 우수한 성능을 보이는 상위 3개 알고리즘 내 순위 변동이 발생했지만, 재최적화 이전과 이후 모두 아동의 내재화 및 외현화 문제행동 예측에 우수한 성능을 보이는 상위 3개 알고리즘의 종류는 변함이 없었다.

학습 데이터에 과도하게 학습되어 구축된 모형이 새로운 데이터에서 정확한 예측을 수행하지 못하는 과적합(overfitting) 문제는 규제화(regularization) 기법을 적용한 별점회귀모형을 활용하여 해결할 수 있으며(노민정·유진은, 2021), 회귀 계수의 가중치에 부여하는 기법 종류에 따라 Ridge, Lasso, Bayesian Ridge 등을 포함한다. 또한 Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor 알고리즘은 대표적인 머신러닝의 앙상블 기법으로, 다수의 개별 모형을 조합하여 최적의 모형을 구축하는 방법이다(신용섭·조형준, 2023). 본 연구에서 아동의 문제행동 예측을 위해 선정된 상위 3개 알고리즘은 별점회귀모형과 앙상블 기법 알고리즘으로 구성되어 있다. 이는 별점회귀모형이나 앙상블 기법은 단일 알고리즘보다 모형 예측력에서 더 높은 성능을 보인다는 기존의 연구와 일치되는 결과이다(정유란·박경원·손수진, 2021; 신용섭·조형준, 2023).

둘째, 최종 선정된 상위 3개 머신러닝 알고리즘을 통해 취학 전 아동의 문제행동 예측에 영향을 주는 주요 변인을 탐색해 보았다. 그 결과, 아동의 학습준비도와 사회적 유능감, 어머니의 양육 스트레스와 우울, 어머니가 지각하는 친가의 사회적 지지는 내재화와 외현화

문제행동 모두에 주요한 영향을 미치는 것으로 나타났으며, 특히 아동의 사회적 유능감, 어머니의 양육 스트레스, 어머니의 우울은 아동의 문제행동에 영향을 끼치는 요인 탐색을 진행한 선행연구와 일치하는 결과이다. 아동의 사회적 유능감의 경우, Campbell(1994)의 연구에서 아동의 개인 내적 변수 중 문제행동에 영향을 주는 주요 변수로 보고된 바 있다. 즉, 사회적 유능감 수준이 낮은 아동은 또래와 상호작용하는 방법에 미숙하거나 자신의 행동 및 정서에 대한 통제가 적절하지 못하는 등의 내재화 및 외현화의 문제행동을 보였다(김민아·이재신, 2004; 이찬숙·현은자, 2008). 또한 다양한 선행연구에서 아동과 가장 많은 정서적 교류를 나누는 어머니가 아동의 문제행동에 밀접한 연관이 있는 대표적인 환경 요인으로 보고되었다(권수진·오새니·이상희, 2016; 문영숙·한진숙, 2002). 그중에서도 어머니의 양육 스트레스와 우울은 양육의 질을 저하시켜 아동의 문제행동에 직·간접적인 영향을 미치는 것으로 나타났다(이영진·노승희, 2017; 하여진, 2018). 더불어 최근 아동의 문제행동에 영향을 주는 개인 외적 요인으로 교사변인의 중요성이 강조되고 있지만(최지은, 2020), 본 연구에서 도출한 주요 변수에 교사 관련 변수는 포함되지 않았다. 이는 취학 전 아동의 내재화 및 외현화 문제행동의 예측에 있어 교사의 영향이 없다는 것이 아니라 아동 개인의 내적 특성과 어머니 관련 변수의 영향력이 상대적으로 더 크다고 해석할 수 있다. 따라서, 앞서 언급한 선행연구와 본 연구결과를 기반으로, 아동의 내재화 및 외현화 문제행동에 공통적으로 밀접한 영향을 미치는 변수는 아동의 개인 내적 요인과 외적 환경 요인 중 특히 어머니 관련 변수인 것으로 확인되었다. 이는 상반되는 특징을 포함하는 내재화 및 외현화 문제행동의 기저에 공유되는 원인이 존재한다는 선행연구(Masten & Curtis, 2000)와 일맥상통하는 결과이다.

셋째, 본 연구 결과를 통해 아동의 내재화 문제행동의 경우 어머니가 평정한 아동의 학습준비도와 어머니가 인식한 친가의 사회적 지지가, 외현화 문제행동의 경우 어머니의 성역할 가치와 어머니가 인식한 친가 및 이웃의 사회적 지지가 기존의 연구에서 다루어지지 않은 새로운 변수로 도출되었다.

먼저, 아동의 내재화 문제행동에서 새롭게 도출된 아동의 학습준비도 변수의 경우, 기존의 선행연구에서 주로 외현화 문제행동과의 관계를 규명하기 위해 활용되었으며(하민경, 2022), 이는 본 연구 결과에서도 외현화 문제행동을 예측하는 주요한 변수 중 하나로 확인되었다. 학습준비도란 학업을 수행할 수 있는 인지적 능력뿐만 아니라 타인과 원활하게 상호작용할 수 있는 사회정서능력을 포함하는 다차원적 개념이다(정진나, 2020; 김진미·홍세영, 2017). 특히 본 연구에서 학습준비도를 측정하기 위하여 사용한 척도는 사회정서발달, 학습에 대한 태도, 의사소통, 인지발달 및 일반적 지식의 네 영역을 포함하고 있다. 이

중 사회정서 발달 영역이 정서적 반응성, 불안/우울 등과 같은 정서적인 측면에서의 행동 문제를 의미하는 내재화 문제행동과도 밀접한 관계를 가질 가능성이 높은 것으로 해석할 수 있다. 또한 내재화 문제행동과 외현화 문제행동은 서로 다른 차원임에도 불구하고 서로 간에 영향을 주고받는 긴밀한 관계성을 가지고 있으며(강지현·김재철·윤경희 외, 2012; 염정원·조한익, 2016), 아동의 인지 및 사회정서의 발달이 제대로 이루어지지 않으면 문제행동이 발생할 수 있다는 선행연구(고인숙·이정숙, 2008; 하민경, 2022) 결과를 미루어 봤을 때, 아동의 언어, 인지, 정서, 사회성의 발달 수준을 포함하는 학습준비도가 아동의 전반적인 문제행동에 영향력을 미친다고 할 수 있다. 따라서 이러한 선행연구와 학습준비도가 외현화뿐만 아니라 내재화 문제행동 예측에 중요한 변수라는 본 연구 결과를 기반으로, 학습준비도가 내재화 및 외현화 문제행동과 구체적으로 어떠한 관계를 맺고 있는지 추가로 탐색해 볼 필요성이 제기된다.

취학 전 아동의 외현화 문제행동 예측에 새롭게 도출된 어머니의 성역할 가치는 해당 변인이 유아의 공격성에 영향력을 미친다는 선행연구(임현주, 2021)와 일치하는 결과이다. 즉 어머니가 인식하는 성역할가치가 양성 간 평등할수록 유아의 공격적 문제행동이 감소하는 경향이 높은 것으로 나타났다. 하지만, 기존의 연구는 성역할가치와 외현화 문제행동의 하위영역인 공격행동과의 관계성만을 살펴보았기에, 본 연구 결과에서와 같이 이러한 결과가 상위영역인 외현화 문제행동과의 관계에서도 확인될 수 있음을 짐작해 볼 수 있다. 더불어, 외현화 문제행동 예측에 미치는 성역할가치의 영향력은 주요 변수로 함께 도출된 공동양육과 연관 지어 생각해 볼 수 있다. 부모가 인식하는 양성적인 성역할가치는 부부간 가사 및 양육 참여가 동등한 수준으로 수행되고 있는지 등으로 평가할 수 있다(이뿐새·류성연·송민영, 2017; 임현주, 2021). 따라서 부모가 지닌 양성평등적 성역할가치가 높을수록 원만한 양육 분담 및 공동양육이 이루어지고(박선주·강민주, 2017), 서로에게 지지적인 부모 공동양육 수준을 높여 유아의 외현화 문제행동을 감소시키는 결과로(유다정·한세영, 2022) 이어진다고 해석할 수 있다. 하지만 부모의 양성평등적 성역할가치와 공평한 공동 분담 수준이 아동의 발달에 미치는 영향을 살펴보는 연구는 부족한 상황이다. 따라서, 성역할가치, 공동양육과 아동의 외현화 문제행동과의 관계를 보다 깊이 있고 면밀하게 탐색하는 후속 연구를 제안한다.

마지막으로, 취학 전 아동의 내재화 및 외현화 문제행동 예측에 주요 변인으로 각각 어머니가 지각하는 친가의 사회적지지와 친가 및 이웃의 사회적지지가 도출되었다. 이는 아동의 문제행동에 영향을 미치는 대표적인 환경요인으로 아동과 많은 정서적 교류를 하는 주양육자인 어머니 관련 변인이 많이 연구되어 왔으며(김정민, 2016; 임정수·이진영·남보

영, 2022; 유다정·한세영, 2022), 그중 어머니의 스트레스 완화와 심리적 안정에 긍정적인 영향을 미치는 사회적지지에 대한 연구가 진행되는 추세이다. 박형원(2011)은 어머니의 높은 외부 접촉과 사회적지지는 아동의 문제행동 수준을 감소시킨다고 하였으며, 용효중(2009)은 다문화가정 어머니가 지각하는 사회적지지 수준이 높을수록 아동의 문제행동이 줄어든다고 하였다. 이와 같은 맥락으로 어머니가 친가나 이웃으로부터 지지받는다고 느낄수록 심리·정서적인 안정감을 얻었으며, 이는 자녀의 문제행동 수준에까지 영향을 미친 것으로 해석할 수 있다. 다만, 김갑순, 박윤조(2016)의 연구에 의하면 어머니에 대한 사회적지지 중 친구 및 동료로부터 받는 사회적 지지가 유아의 내재화 및 외현화 문제행동에 영향을 미치는 것으로 나타나, 본 연구결과와 차이가 났다. 따라서 어머니가 지각하는 사회적지지의 여러 유형과 아동의 내재화 및 외현화 문제행동 사이 관계에 관한 다양한 연구가 진행되어야 할 것이다.

본 연구는 머신러닝의 다양한 회귀 모델 알고리즘을 활용하여 아동/부모/교사/육아기관 변수를 동시에 투입함으로써, 취학 전 아동의 내재화 및 외현화 문제행동을 예측하는 변수들을 총괄적으로 탐색하고자 하였다. 즉, 기존 연구에서 활용된 통계 분석 방법이 보였던 변수 투입 및 선택에서의 한계를 극복하고, 여러 개의 변수를 투입함으로써 내재화 및 외현화 문제행동 예측에 중요한 개인 및 환경요인을 한 번에 도출할 수 있었다. 특히 선행연구에서 다루어지지 않았던 새로운 변수를 도출해냄으로써, 아동의 내재화 및 외현화 문제행동에 대한 모형 및 이론을 확장하고 더 깊이 있게 이해할 수 있는 기초 정보를 제공하였다는 점에서 학술적 의의를 찾을 수 있을 것이다. 또한 본 연구에서 새롭게 도출된 주요 변수들은 완전히 새로운 개념이 아닌 기존의 선행연구에서 언급된 설명변수들과 어느 정도 연관성을 가지고 있는 것으로 나타났다. 따라서 본 연구결과를 통해 새롭게 도출된 변수들과 기존의 선행연구에서 언급된 변수들과의 관계성에 대해 더욱 세밀한 연구가 이루어질 수 있는 발판을 제공하였다는 의의가 있다.

이러한 의의에도 불구하고 본 연구는 몇 가지 제한점을 가지고 있으며, 이를 바탕으로 후속 연구를 위한 제언을 하고자 한다. 첫째, 아동의 문제행동을 예측하는 다양한 회귀모형 알고리즘의 성능을 여러 평가 지수를 기반으로 평가하였으나, 평가 지수 중 하나인  $R^2$  값이 전반적으로 낮아 예측력이 높지 않았다. 다만, 해당  $R^2$ 의 값은 분석에 사용된 140개의 설명변수 모두를 투입한 모형의 예측력으로, 다수의 설명변수의 투입으로 인한 모형의 복잡도 증가에 따른 결과로 해석할 수 있다(류나현·김형석·강필성, 2016). 따라서 본 연구에서 도출된 주요 변수만을 대상으로 모형을 설정하여 모형의 예측력에 대해 재검증을 수행하는 추후 연구가 필요하다고 판단된다. 둘째, 본 연구에서는 아동의 문제행동 예측에

미치는 변인의 영향력 크기를 기반으로 변수의 중요도를 산출하였다. 하지만 본 연구에서 사용한 변인의 영향력을 나타내는 값은 영향력의 크기에 대한 정보를 줄 뿐, 영향력의 통계적 유의성 또는 그의 방향성을 확인하기는 어렵다. 따라서 후속 연구를 통해 본 연구에서 도출된 영향력이 큰 변수들을 대상으로 아동의 문제행동에 대해 통계적으로 유의한 지, 유의하다면 그 방향성이 긍정적인지 부정적인지 등을 탐색하여 통계적 추론까지 고려할 수 있는 연구를 설계할 것을 제안한다. 셋째, 영유아기 내재화 및 외현화 문제행동에 유의미한 영향을 미치는 변인들 중 본 연구에서 사용된 패널자료에 구축되지 않은 변인은 설명변수로 투입하지 못했다는 제한점이 존재한다. 이러한 연구의 제한점을 극복하기 위하여 본 연구 결과를 기초 자료로, 다양한 패널자료를 활용하여 아동의 문제행동에 대한 면밀하고 세밀한 후속 연구 진행을 제안하는 바이다.

## V. 정책적 제언 및 시사점

본 연구의 결과를 바탕으로 제시하는 정책적 제언과 시사점은 다음과 같다.

첫째, 취학 전 아동의 문제행동을 유발할 수 있는 원인을 사전에 방지하기 위해 부모에 대한 지원을 강화할 필요성이 있다. 앞서 언급한 바와 같이, 최근 문제행동을 보이는 아동의 비율과 아동이 경험하는 문제행동의 종류가 증가하고 있는 추세이다(성다겸·김춘경, 2017; 이운경·김민주·윤기봉, 2018). 2005년 서울시에서 소아청소년을 대상으로 실시한 정신 질환 유병률 조사 결과에서 약 25%가 정신건강 문제를 가지고 있는 것으로 나타났다(강은정, 2007). 또한 2016년부터 2020년까지 지난 5년간 우울 및 불안장애 등의 진료를 받은 0~9세까지 아동이 꾸준히 증가하고 있으며, 특히 기타 행동 및 정서장애가 상위를 차지하고 있었다(박진우·허민숙, 2021). 이러한 문제행동 증가와 그 지속성으로 인해 유아의 문제행동에 대한 조기 발견 및 개입과 예방의 중요성이 대두되고 있다. 특히 유아기에 발현되는 문제행동은 신체적 및 심리·정서적 발달에 부정적 영향을 미치며, 이후 청소년기의 비행이나 일탈과 같은 심각한 사회적 부적응 행동으로 이어질 수 있다(우수경, 2017). 유아기는 아동기나 청소년기, 성인기에 비하여 문제행동의 개선 및 안정화의 가능성이 높아(지성애·김성현, 2014) 이와 관련된 지원이 중요하지만, 2015년에 수립되고 2022년에 개정된 아동복지법 중 문제행동과 관련된 지원정책은 부족한 것으로 확인된다(하여진, 2018). 이에, 여성가족부에서 지원하는 다양한 부모역할지원 사업에 부모가 자녀의 문제행동을 간단하게 진단할 수 있는 자료와 유아기에 발병률이 높은 문제행동에 대한 프로그램

내용을 추가하는 방안을 고려해 볼 수 있을 것이다. 더불어, 문제행동 수준이 높다고 진단된 유아에 대해서는 전문 상담기관과 연계하여 부모와 유아교육 기관 교사가 아동의 문제행동을 지속적으로 살펴보고 개선할 수 있도록 하는 지원정책도 필요할 것으로 간주된다.

둘째, 취학 전 아동의 부모가 자녀의 문제행동을 예방하기 위하여 양성평등적 성역할가치를 바탕으로 공동양육을 실천할 수 있도록 하는 정책이 필요한 것으로 보인다. 현대사회의 급격히 추진된 산업화 및 도시화를 통해 여성의 경제활동 참여가 증가하였고, 이에 따라 전통적으로 가정 내에서 구분되어 왔던 성역할의 범위가 변화되고 있다. 이 중 부모가 지각하는 성역할가치는 부부간 공동양육과 함께 고려해야 할 변인으로, 부모가 성역할에 대한 고정관념을 가지지 않고 평등한 관점을 가질수록 가정 내 가사 및 양육에 대한 부담이 원만하게 이루어지는 것으로 나타났다(박선주·강민주, 2017). 특히 오민아, 변길진, 권정윤(2018)에 따르면 아버지가 높은 수준의 양성 평등적 성역할가치를 가질수록 공동양육에 더 협조적으로 참여하는 것으로 확인되어, 어머니뿐만 아니라 아버지가 지닌 성역할가치가 자녀양육 및 건강한 가정형성에 중요하게 작용하고 있음을 알 수 있다. 최근 여성의 사회진출로 인해 부부의 공동양육 참여에 대한 필요성이 증가함에 따라, 정부는 2023년 ‘부모 맞돌봄 문화’를 활성화하기 위한 부모의 공동육아 인센티브를 확대 및 개편하고 남성 육아 참여 확대 및 여성의 육아 부담 완화를 위한 정책 방안을 발표하였다(정책브리핑, 2023. 1. 16). 하지만 해당 정책에는 부모의 공동양육 및 아버지의 육아 참여 증대를 위한 양적인 부분만 포함되어 있을 뿐 질적인 측면에 대한 노력이 부족한 것으로 판단된다. 따라서 전통적인 성역할 가치관에서 벗어나 평등적인 가치관을 바탕으로 부모가 적극적으로 공동양육에 참여할 수 있도록 ‘부모 맞돌봄 문화’ 정책을 시행하는 데 양성 평등적인 부모역할훈련 교육 프로그램을 기획하여 추가하는 방안을 고려해 볼 것을 제안한다.

## 참고문헌

- 강은정(2007). 한국 아동 정신건강 현황과 정책과제. 보건복지포럼, 128, 60-72.
- 강지현·김재철·윤경희·임은미(2012). 초등학생의 내재화-외현화 문제 잠재계층에 대한 다문화, 성, 부모, 개인역량의 효과. 교육종합연구, 10(4), 201-224.
- 고은경(2020). 잠재계층성장분석과 머신러닝에 의한 초등학교 저학년 학업수행능력의 종단 유형 예측. 학습자중심교과교육연구, 20(18), 1211-1230.
- 고인숙·이정숙(2008). 유아의 사회적 유능감 및 정서지능이 문제행동에 미치는 영향. 한국아동심리치료학회지, 3(1), 93-111.

- 권수진·오새니·이상희(2016). 유아의 문제행동 관련변인 연구: 유아변인, 어머니변인, 또래 관계변인. 미래유아교육학회지, 23(4), 385-406.
- 김갑순·박윤조(2016). 어머니에 대한 사회적 지지와 어머니와 자녀의 상호작용이 유아의 문제행동에 미치는 영향. 한국보육학회, 16(3), 27-44.
- 김민아·이재신(2004). 어머니 특성 및 유아의 기질과 유아의 부적응 행동과의 관계에 대한 연구. 유아교육연구, 24(5), 145-166.
- 김수정·정익중(2015). 가정환경이 유아발달에 미치는 영향과 교사효능감의 조절효과. 육아정책연구, 9(1), 1-28.
- 김연·한태숙·황혜정(2004). 유아의 정서지능, 인지능력, 기질과 문제행동의 관계성 및 영향력에 관한 연구. 유아교육연구, 24(6), 277-298.
- 김영미·송하나(2018). 유아기 자녀를 둔 아버지 우울, 온정적 양육행동, 유아의 외현화 문제행동 간의 자기회귀교차지연 효과 검증. 아동학회지, 39(3), 73-82.
- 김영심·전정민·김정미(2020). 부모의 정서표현, 정서표현 양가성이 유아의 문제행동에 미치는 영향. 미래유아교육학회지, 27(3), 59-82.
- 김영아·이진·문수중·김유진·오경자(2009). 한국판 CBCL 1.5-5의 표준화 연구. 한국심리학회지: 임상, 28(1), 117-136.
- 김정민(2016). 어머니의 우울과 양육효능감이 아동의 문제행동에 미치는 영향. 사회과학논총, 15, 91-104.
- 김지혜(2011). 어머니와 교사가 인식한 유아의 행복과 유치원 적응 및 문제행동과의 관계. 동의대학교 석사학위논문.
- 김진미·홍세영(2017). 유아의 학교준비도 영향요인 분석. 육아정책연구, 11(2), 169-194.
- 김태련·강우선·김경은·김도연·김문주·박랑규·서수정·양혜영·이경숙·장은진·조현섭·허묘연(2003). 발달장애 심리학. 서울: 학지사.
- 김효정·조동혁·조희준(2023). 일-가정 양립 갈등이 부모역할만족도와 유아 문제행동에 미치는 영향에 관한 연구. 학습자중심교과교육연구, 23(6), 195-209.
- 노민정·유진은(2021). 별점회귀모형에서의 통계적 추론: 청소년의 스마트폰 의존 관련 변수를 중심으로. 한국청소년연구, 32(1), 147-174.
- 류나현·김형석·강필성(2016). 다중선형회귀모형에서의 변수선택기법 평가. 대한산업공학회지, 42(5), 314-326.
- 문영숙·한진숙(2002). 어머니의 자녀양육행동이 아동행동문제에 미치는 영향. 부모자녀건강학회, 5(2), 161-176.
- 문용혁·신익희·이용주·민옥기(2019). 자동 기계학습 (AutoML) 기술 동향. 전자통신동향분석, 34(4), 32-42.

- 박선주·강민주(2017). 맞벌이 가정 아버지의 성역할태도와 어머니 문지기 역할이 아버지의 양육 참여도에 미치는 영향. *육아정책연구*, 11(3), 87-111.
- 박세준·변영철(2021). 전동 모빌리티 수요 예측 정확도 개선. 2021 한국정보기술학회 추계 종합학술대회 논문집, 104-106.
- 박소영·정혜원(2020). 중학생의 진로 결정 예측변수 탐색: 머신러닝 기법 적용. *아시아교육연구*, 21(3), 727-753.
- 박진우·허민숙(2021). 아동청소년의 정신건강 현황, 지원제도 및 개선방향. 서울: 국회입법조사처.
- 박형원(2011). 아동의 행동문제에 영향을 미치는 어머니 관련 변인에 관한 연구. *아동과 권리*, 15(2), 205-225.
- 서석원(2019). 어머니의 자아존중감과 양육스트레스가 유아의 문제행동에 미치는 영향. *생태 유아교육연구*, 18(4), 51-71.
- 서석원(2022). 부모의 행복감이 유아의 사회적 유능성을 매개로 유아의 내재화 문제행동에 미치는 영향. *한국산학기술학회 논문지*, 23(8), 300-309.
- 성다점·김춘경(2017). 어머니의 출산 전 우울과 임신 시 느낌이 유아의 내재화 및 외현화 문제행동에 미치는 영향: 부모교육 경험유무를 중심으로. *놀이치료연구*, 21(2), 19-33.
- 송은정·김지윤·안재별·홍세희(2022). 머신러닝을 활용한 초등학생의 자아탄력성 영향 변수 및 예측모형 탐색. *아동교육*, 31(3), 105-122.
- 신소정·이재모(2011). 지역사회 교정을 통한 비행 청소년 재범 방지에 관한 연구. *교정복지연구*, 53-90.
- 신웅섭·조형준(2023). 양상불 비모수적 변수 선택 방법. *한국자료분석학회*, 25(2), 523-536.
- 신유림(2021). 교사-유아관계 군집유형별 유아의 성, 기질 및 문제행동에서의 차이. *생태 유아교육연구*, 20(4), 1-21.
- 원영미·박혜원·이귀옥(2004). 유아의 기질과 가정환경 및 유아의 부적응 행동 간의 관계: 연변조선족과 한국 유아의 비교연구. *유아교육연구*, 24(2), 331-334.
- 연은모·윤해옥·최효식(2016). 부모의 우울, 지각된 가족기능, 양육태도, 유아의 내재화 및 외현화 문제행동 간의 구조적 관계 분석: 자기효과와 상대방효과를 중심으로. *유아교육연구*, 36(2), 243-269.
- 염수진·이희영·최태진(2022). 유아교사의 문제행동지도 효능감에 따른 유아의 문제행동 및 사회적 능력. *인문사회과학연구*, 23(3), 167-200.
- 염정원·조한익(2016). 청소년의 외현화, 내재화 문제행동과 학교생활적응의 종단적 인과관계 분석. *교육심리연구*, 30(1), 195-223.
- 오민아·변길진·권정윤(2018). 유아기 자녀를 둔 아버지 성역할 가치와 양육분담이 어머니의

- 양육스트레스와 일-가정 양립 어려움에 미치는 영향: 아버지의 양육스트레스를 매개로. 유아교육학논집, 22(5), 429-448.
- 용효중(2009). 결혼이민여성의 문화적응 및 가족관계 스트레스, 사회적 지지가 아동기 자녀의 정신건강에 미치는 영향에 관한 연구. 가톨릭대학교 대학원 석사학위논문.
- 우수경(2017). 유아 문제행동의 변화양상 및 영향요인 분석. 열린유아교육연구, 22(5), 381-400.
- 유다정·한세영(2022). 부모공동양육과 유아의 정서 조절이 유아의 외현화 문제행동에 미치는 영향. 아동학회지, 43(2), 95-110.
- 유진은·김형관·노민정(2020). Group Mnet 기계학습 기법을 통한 중학생의 끈기( grit) 관련 변수 탐색. 한국청소년연구, 31(1), 157-182.
- 이뿐새·류성연·송민영(2017). 아버지의 성역할가치가 어머니의 양육스트레스와 양육효능감에 미치는 영향: 어머니 자아존중감의 매개효과. 유아교육연구, 37(6), 75-97.
- 이성봉·방명애·김은경·박지연(2014). 정서 및 행동장애. 서울: 학지사.
- 이영진·노승희(2017). 부모의 양육스트레스와 우울이 유아의 문제행동에 미치는 영향. 유아교육학논집, 21(4), 301-324.
- 이운경·김민주·윤기봉(2018). 맞벌이 부모의 일-가정 갈등, 온정적 양육행동 및 학령초기 아동의 문제행동 간의 관계. 아동학회지, 39(3), 141-156.
- 이은주·이기연(1995). 아동의 과외수업과 문제행동의 관계에 관한 연구: 국민학교 저학년아동을 대상으로. 한국아동복지학, (3), 205-239.
- 이지영·이상희(2019). 유아개인변인, 또래놀이행동, 사회적유능감, 문제행동이 초등학교 적응에 미치는 영향. 한국유아교육·보육복지연구, 23(3), 171-199.
- 이지예·신경순·김영희(2017). 아버지의 심리적 스트레스와 양육스트레스 및 양육행동이 유아의 문제행동에 미치는 영향. 영유아아동정신건강연구, 10(1), 59.
- 이진숙·한지현·박혜원(2016). 유아의 기질과 지능, 문제행동의 관계 분석. 아동학회지, 37(5), 117-128.
- 이찬숙·현은자(2008). 유아의 기질, 사회적 유능감, 감성지능, 도덕성 및 부모양육태도와 유아의 문제행동간의 관계. 아동학회지, 29(3), 223-238.
- 임정수·이진영·남보영(2022). 아버지의 양육태도와 아동의 내재화 및 외현화 문제행동의 관계에서 어머니 양육스트레스의 매개효과. 아동복지연구, 20(2), 67-93.
- 임현주(2015). 유아의 성별에 따른 내재화, 외현화 문제행동에 영향을 미치는 변인에 관한 연구. 유아교육연구, 35(2), 97.
- 임현주(2021). 어머니의 양육효능감, 공동양육, 성역할 가치가 유아의 공격성에 미치는 영향. 인문사회21, 12(3), 1721-1734.
- 정미라·이정은·박혜성(2013). 유아의 정서 및 행동문제 선별도구로서 집-나무-사람 그림 검

- 사의 활용가능성. 유아교육연구, 33( 6), 367- 386.
- 정유란·박경원·손수진(2021). 계절내 예측 일 최고기온의 예측 성능 향상을 위한 심층학습 앙상블 알고리즘. 한국기상학회 학술대회 논문집, 2021(10), 349-349.
- 정진나(2020). 유아의 외현화 문제행동과 학교준비도의 관계에서 또래 놀이 상호작용의 중단매개효과 검증. 인지발달장애학회지, 11(4), 89-106.
- 정책브리핑(2023. 1. 16). 맞돌봄문화 확산, 여성경력단절 최소화하도록 육아휴직 기간 연장 안 마련하겠다. <https://www.korea.kr/news/policyNewsView.do?newsId=148910635> 에서 2023년 5월 22일 인출함.
- 정혜원·김예림·박소영(2020). 초·중학생의 그림에 영향을 미치는 변수 탐색: 별점화 회귀모형 적용. 학습자중심교과교육연구, 20(8), 673-693.
- 조정숙(2015). 정서행동문제 아동을 위한 통합치료의 이론과 실제. 서울: 신아사.
- 지성애·김성현(2014). 유아의 행동문제와 인지능력 및 자아존중감 간의 관계분석. 유아교육연구, 34(3), 197-219.
- 최선녀·최항준(2017). 유아의 일상적 스트레스가 문제행동에 미치는 영향. 한국산학기술학회 논문지, 18(3), 628-639.
- 최은정·김은향(2019). 가족상호작용, 부모-자녀상호작용, 놀이상호작용이 유아의 문제행동에 미치는 영향: 사회적 유능감의 매개효과를 중심으로. 열린유아교육연구, 24(2), 1-21.
- 최지은(2020). 교수효능감, 유아 사회적 유능감, 유아 내재화 문제행동, 유아 외현화 문제행동 간의 구조적 관계 분석. 미래유아교육학회지, 27(4), 1-24.
- 하민경(2022). 유아 학습준비도, 부부갈등, 어머니 우울이 유아 외현화 문제행동에 미치는 영향. 인문사회 21, 13(1), 2453.
- 하여진(2018). 유아의 문제행동 양상에 따른 잠재계층분류와 영향요인 검증. 육아정책연구, 12(1), 125-149.
- 하영례·정효은(2009). 기관변인에 따른 5세 유아의 문제행동. 한국영유아보육학, 57, 171-191.
- Achenbach. T. M., & Rescorla. L. A. (2000). Manual for ASEBA preschool forms & profiles. Burlington, VT: Research Center for Children, Youth, and Families, University of Vermont.
- Ali, M. (2020). PyCaret: An open source, low-code Machine Learning library in Python. <https://pycaret.org/> (accessed February 2023)
- Campbell, S. B. (1994). Hard-to-manage preschool boys: Externalizing behavior, social competence, and family context at two-year follow up. *Journal of Abnormal Child Psychology*, 22(2), 147-166.

- Dangeti, P. (2017). *Statistics for machine learning*. Packt Publishing Ltd.
- Gain, U., & Hotti, V. (2021). Low-code AutoML-augmented data pipeline—a review and experiments. In *Journal of Physics: Conference Series*, 1828, 012015.
- Garber, J. (1984). Classification of childhood psychopathology: A developmental perspective. *Child development*, 55(1), 30-48.
- Greenberg, M. T., Speltz, M. L., & Deklyen, M. (1993). The role of attachment in the early development of disruptive behavior problems. *Development and psychopathology*, 5(1-2), 191-213.
- Kaiser, B., & Rasminsky, J. S. (2007). *Challenging behavior in young children*. New York, NY: Pearson.
- Masten, A. S., & Curtis, W. J. (2000). Integrating competence and psychopathology: Pathways toward a comprehensive science of adaptation in development. *Development and Psychopathology*, 12, 529-550.
- Poulou, M. S. (2015). Emotional and behavioural difficulties in preschool. *Journal of Child and Family Studies*, 24, 225-236.

- 논문접수 8월 3일 / 수정본 접수 9월 13일 / 게재 승인 9월 20일
- 교신저자: 홍세희, 고려대학교 교육학과 교수, seheehong@korea.ac.kr

## Abstract

### Exploring the Factors Influencing Internalizing and Externalizing Problem Behaviors in Preschool Children Using Machine Learning Methods

JiYoon Kim, An Saebuy, Sehee Hong

This study aimed to explore the factors that influence internalizing and externalizing problem behaviors in preschool children and to examine these factors using machine learning methods with the data from the 7th year of PSKC. Based on a result of analyses using top regressor algorithms, learning readiness and paternal social support for internalizing problem behavior and the value of mothers' gender roles and paternal and neighbor social support for externalizing problem behavior were newly derived as major factors. The results suggest that the factors influencing problem behaviors can be identified, thereby enabling the establishment of educational policies and strategies to prevent problem behaviors.

Key words: Problem behavior, Internalizing and externalizing problem behavior, Preschool children, Machine learning