

## 랜덤포레스트 머신러닝 기법을 활용한 전통적 비행이론기반 청소년 온·오프라인 비행 예측요인 연구\*

이택호<sup>1)</sup> 김선영<sup>2)</sup> 한윤선<sup>†</sup>

본 연구에서는 청소년 비행이 지속적인 사회문제로 대두됨에 따라 청소년의 온·오프라인 비행을 예측하는 주요 요인들을 탐색하고 전통적 비행이론(사회학습이론, 일반긴장이론, 사회통제이론, 일상활동이론, 낙인이론)의 적용 가능성을 살펴보았다. 분석에 활용된 데이터는 한국아동·청소년패널조사 2010(KCYPS 2010)의 초1, 초4, 중1 패널 6차년도 데이터이다(N=4,137). 예측 모형을 구축함에 있어 전통적 통계기반의 회귀모형 대신 랜덤포레스트 머신러닝 기법을 활용함으로써 예측 성능 향상과 더불어 보다 많은 예측요인의 고려 가능성에 초점을 두었다. 랜덤포레스트 분석 결과, 청소년의 온·오프라인 비행을 설명하는 데에 전통적인 비행이론은 여전히 유효하였으며, 온라인 비행은 주로 개인적 요인(일상활동이론, 낙인이론)과, 오프라인 비행은 사회적 요인(사회학습이론, 사회통제이론)과 관련이 있는 것으로 나타났다. 또한 일반긴장이론은 온라인 비행과 오프라인 비행 모두를 예측하는 중요한 이론적 기반임을 확인할 수 있었다. 본 연구는 머신러닝 기법을 통해 청소년 비행에 영향을 주는 주요 요인을 도출하고, 기존 비행이론의 활용 가능성도 함께 고려했다는 점에서 의의가 있으며, 청소년 온·오프라인 비행에 대한 예방 및 개입 방향성을 제고하는 기반을 제공할 것이라 기대된다.

주요어 : 청소년 비행, 온라인 비행, 오프라인 비행, 랜덤포레스트, 비행 이론

\* 이 논문은 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020S1A5A2A03045529).

1) 서울대학교 사회복지학과 박사과정

2) 서울대학교 사회복지학과 석사졸업

† 교신저자 : 한윤선, 서울대학교 사회복지학과 부교수, (08826) 서울특별시 관악구 관악로 1,

E-mail: yshan@snu.ac.kr



Copyright ©2022, The Korean Psychological Association of Culture and Social Issues  
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

청소년의 비행은 청소년에게 기대되는 사회적 규범에서 벗어난 일탈 행위를 의미한다(김준호 등, 2013). 국내에서 청소년 비행은 여전히 심각한 사회문제로 대두되고 있으며, 온라인 활동이 청소년의 일상에서 큰 비중을 차지함에 따라, 비행의 영역이 오프라인을 넘어 온라인으로 확장되는 양상으로 전개되고 있다. 특히 국내 관련 자료를 살펴보면 청소년 비행의 저연령화와 사이버 범죄의 증가 양상이 두드러짐을 알 수 있다. 예컨대, 법무연수원(2022)에서 제공하는 2021 범죄백서에 따르면 전체 범죄자 중 소년범죄자의 점유비율이 2015년부터 2020년까지 지속적으로 증가하는 경향성을 보였다. 특히 살인, 강도, 폭력과 같은 강력범죄와 절도, 사기, 횡령과 같은 재산범죄에 대하여 16세~18세의 범죄 비율은 일정 수준을 유지하거나 감소세를 보였으나, 14세~15세는 2015년 이후 전반적으로 증가한 것으로 나타났다. 경찰청에서 제공하는 경찰통계연보에서도, 10대 청소년의 사이버 범죄 검거 인원이 2018년 8,642명, 2019년 9,651명, 2020년 12,165명으로 꾸준히 증가한 것을 확인할 수 있다. 특히, 정보통신망 이용 범죄인 사이버사기, 사이버 금융 범죄 등은 2018년 6,713건에서 2020년에 9,130건으로 크게 늘어났으며, 사이버 음란물, 사이버 도박 등 불법콘텐츠 범죄도 2018년 1,793건에서 2020년 2,870건으로 대폭 증가하였다(경찰청, 2019; 경찰청, 2020; 경찰청, 2021). 이처럼 법률에 저촉되어 수면위로 드러나는 청소년의 비행만 살펴보다도 그 심각성에 대하여 문제의식을 갖고 주시해야 할 필요성이 있다.

앞서 언급한 사이버 범죄의 증가와 더불어 법적 영향력 밖에 있는 온라인 비행의 심각성 또한 사회적인 관심이 필요하다. 이와 관련하여

여, 2020년 청소년 매체 이용 및 유해환경 실태조사(여성가족부, 2020)에 따르면, 청소년의 90.4%가 주 1회 이상 모바일 메신저를 이용하고, 90.7%가 실시간 방송 및 동영상 시청하며, 65.0%의 청소년이 SNS를 이용하고 있는 것으로 나타났다. 온라인 활동과 관련된 역기능 중 하나인 스마트폰 과의존 또한 전 연령대 중 청소년에게서 위험군의 비율이 가장 높았다(과학기술정보통신부, 2021). 이처럼 온라인 활동을 위한 매체 이용률이 높고, 해당 매체에 영향을 받기 쉬운 청소년들은 온라인 비행에 노출될 가능성이 크다. 실제로 2021년에 조사된 사이버폭력 실태조사(방송통신위원회, 2021)에서 청소년의 사이버폭력<sup>1)</sup> 가해 및 피해 경험률이 각각 14.1%와 23.4%로 모두 성인에 비해 높음을 보고하고 있다. 앞선 논의를 종합해보면 현대의 청소년 비행을 효율적으로 예방하고 개입하기 위해서는 오프라인을 비롯해 온라인 영역을 함께 고려할 필요가 있다. 그 노력의 일환으로 본 연구에서는 청소년의 온·오프라인 비행을 유발하는 심리·사회적 요인을 종합적으로 탐색하고 실천적 개입 방향을 제시하고자 한다.

청소년의 비행을 유발하는 요인은 여러 전통적인 비행이론(일반긴장이론, 사회통제이론, 사회학습이론, 일상활동이론, 낙인이론 등)에서 제시된 바 있다. 하지만 이론적 바탕이 되는 사회 현상들은 이론이 등장한 20세기 중후반의 오프라인 비행에 대부분 한정되어 있다.

1) 사이버폭력 실태조사(방송통신위원회, 2021)에서 사이버폭력은 ‘사이버 언어폭력’, ‘사이버 명예훼손’, ‘사이버 스토킹’, ‘사이버 성폭력’, ‘신상정보 유출’, ‘사이버 따돌림’ 등 사이버(인터넷, 휴대전화 등) 공간에서 언어, 영상 등을 통해 타인에게 피해 혹은 불쾌감을 주는 행위로 정의됨.

따라서, 해당 이론들을 바탕으로 이후에 등장한 온라인 비행을 설명하기 위해서는 기존 이론에서 제시한 비행 위험 요인들이 설명력을 갖는지 확인해 볼 필요가 있다. 선행연구에서는 특정 이론이 제시한 비행 관련 요인들을 바탕으로 오프라인 혹은 온라인 비행을 예측하는 시도가 있었다(이응택, 이은경, 2021; 정혜원, 2009; 박영신, 김의철, 2010; 서경현, 2002). 하지만, 여러 이론에서 제시하는 변인들을 통합적으로 고려해 청소년의 온·오프라인 비행을 예측하는 연구는 상대적으로 부족하다.

선행연구에서의 이러한 한계점은 그동안 주로 사용된 분석방법론의 한계와 관련이 있다. 청소년 비행 관련 실증연구들의 대부분은 전통적 통계기반의 회귀모형을 적용하였다. 회귀모형이 갖는 장점들이 있으나, 다음과 같은 한계점들 또한 존재한다. 첫째로, 확률분포를 기반으로 모수를 추정하는 회귀모형은 독립변수의 외생성(exogeneity), 오차항의 분포가정 등과 같은 다양한 가정들을 전제하고 있지만, 실제 분석 과정에서 해당 가정들이 성립되지 못하는 경우가 많다(최필선, 민인식, 2018). 둘째로, 다중공선성 문제로 인해 한 모형에 많은 변수를 투입할 수 없는 경우가 빈번하다. 이러한 전통적 통계 모형의 제한점들을 보완하고자 최근 국내 사회과학 분야에서는 머신러닝 기법을 도입하는 움직임이 나타나고 있다(이은정, 조희숙, 소영수, 2020). 다양한 머신러닝 기법 중 랜덤포레스트 기법은 사회과학 연구에 적합한 특징들을 지닌 방법론으로 알려져 있다(박소영, 정혜원, 2020; 김관준, 2019).

앞선 논의를 바탕으로, 본 연구는 랜덤포레스트 기법을 활용해 일반긴장이론, 사회통제

이론, 사회학습이론, 일상활동이론, 낙인이론 등 다양한 이론에 기반한 비행 관련 요인들을 예측모형에 동시 투입함으로써 비행 예측 성능을 높이고자 하였다. 나아가 전통적인 이론들에서 제시하는 요인들이 청소년 비행 예측모형에서 각각 어떠한 예측력을 나타내는지 탐색하고, 온라인과 오프라인 비행에 따라 예측 요인에 어떠한 차이가 나타나는지 살펴보고자 하였다. 즉, 본 연구는 전술한 실천적, 방법론적, 이론적 필요성을 근거로 온·오프라인 비행을 예측하는 머신러닝 모형을 통해 주요한 심리·사회적 요인을 비교·대조하고 과거로부터 정립된 청소년 비행 이론의 온·오프라인 비행에 대한 적용 가능성을 검토할 것이다. 그리고 이를 통해 지속적으로 심각성을 더해가는 청소년 비행에 대한 예방 및 개입 방향성을 재고하는데 필요한 연구 기반을 제공하고자 한다.

#### 청소년의 온·오프라인 비행 관련 선행연구 고찰

비행의 개념과 범위는 청소년을 둘러싼 사회적·문화적 상황에 따라 변화할 수 있다(이종원, 이순래, 정윤미, 2016). 정보통신기기의 대중화 및 정보화가 이루어짐에 따라 비행의 범주 또한 오프라인 비행에서 온라인 비행으로 확장되었다. 온라인 비행은 온라인 공간상의 일탈 행위로 컴퓨터, 스마트폰 등 다양한 정보통신매체를 활용하여 이루어진다(이석영, 이택호, 한윤선, 2015; 홍정윤, 이수정, 2020).

선행 연구들은 다양한 방식으로 비행의 유형을 구분해왔으며, 대표적인 것이 비행의 심각성에 따라 경비행과 중비행으로 분류하는 것이다(황성현, 2010). 경비행은 비교적 심각도

가 낮으나 청소년의 사회적 지위에 부적절한 일탈행동으로 흡연, 음주, 가출, 무단결석, 성관계 등이 포함된다(김준범, 장대연, 최장원, 서정열, 2018). 중비행의 경우, 폭력, 절도, 살인, 강간 등 사회적으로 큰 문제를 초래하며 형벌 법령에 위배되는 행위들을 포함한다(김준범 등, 2018; 황성현, 2010). 이러한 구분 방식은 전통적인 비행인 오프라인 비행을 기준으로 확립되었으나, 온라인 비행 또한 청소년의 지위에 맞지 않는 일탈행동부터 법률에 저촉되는 범죄행위까지 심각성에 따라 다양한 비행이 존재한다(박선수, 2018).

비행을 분류하는 또 다른 대표적인 방식은 비행 행위의 결과에 따라 지위비행, 폭력비행, 재산비행으로 분류하는 것이다(임미령, 김지훈, 김미희, 이현, 정애경, 2015). 먼저 지위비행은 경비행과 유사하게 청소년의 지위에 어긋나는 행위를 의미한다. 오프라인 공간에서 발생하는 지위비행으로는 음주, 흡연, 성관계 등이 대표적이며 지위비행에 해당하는 온라인 비행에는 성인 만화, 도박, 게임 등이 있다(양준영, 김지원, 김수영, 홍세희, 2019). 폭력비행은 타인에게 의도적으로 신체적, 정신적, 성적 피해를 주는 것을 의미한다. 오프라인 공간에서 발생하는 폭력비행에는 신체폭행, 또래괴롭힘, 집단폭력, 성폭력 등이 있으며 온라인에서의 욕설, 폭언, 해킹 등이 온라인에서 발생하는 폭력비행에 해당한다(이석영 등, 2015). 재산비행은 타인의 재산을 갈취하거나 훼손하는 행위로(김준호 등, 2013) 오프라인 비행의 예시로는 절도, 강도, 사기 등이 있다(박지수, 김하영, 유진경, 한운선, 2017; 임미령 등, 2015). 온라인에서는 저작권 침해, 타인의 개인정보를 도용해 물적/지적 재산을 훔치는 행위, 온라인 사기거래 등의 형태로 재산비행이 발생

한다. 앞선 예시를 통해 각 공간에서 일어나는 비행에 차이가 존재하지만, 비행을 분류하는 기준은 온라인과 오프라인을 막론하고 적용 가능함을 알 수 있다.

한편, 온라인 공간이 지닌 특성은 온라인 비행을 전통적인 오프라인 비행과 구분되게 한다. 온라인 공간을 오프라인 공간과 차별되게 하는 대표적인 특성들로 크게 익명성, 비대면성, 시공간의 초월성, 피해의 과대성/광역성, 영구성을 들 수 있다(이성식, 2008; 홍정운, 이수정, 2020). 오프라인 공간에서와 달리 온라인 공간에서는 익명성이 보장되기 때문에, 청소년들은 자신의 정보를 완전히 숨기는 것이 가능하다. 이는 탈억제 효과를 유발함과 동시에 비행이 발각될 위험을 낮추어 청소년의 비행 가능성을 높인다(홍정운, 이수정, 2020). 또한, 얼굴을 마주하지 않고 비행을 저지를 수 있는 비대면성으로 인해 죄의식이 줄어들고 청소년은 욕설, 성희롱 등 부적절한 행위에 더욱 대담해진다(양준영 등, 2019). 오프라인 공간과 구분되는 또 다른 특징은 온라인 공간에서는 시공간적 제약이 존재하지 않는다는 점이다. 온라인에 접속만 가능한 상황이라면 청소년들은 언제든지 비행에 가담할 수 있으며, 물리적 거리와 관계없이 누구에게나 피해를 줄 수 있다. 이러한 물리적 제약의 부재는 온라인 비행의 과대성/광역성으로 이어진다. 온라인 공간에서 청소년은 많은 사용자를 대상으로 비행을 저지를 수 있으며, 피해 규모는 빠른 속도로 확장된다. 명예훼손과 같이 피해자에 대한 잘못된 정보가 급속도로 퍼지는 것이 과대성과 광역성의 대표적인 예시이다. 비행 피해가 영구적으로 남아 피해자의 고통을 가중할 수 있다는 점 또한 온라인 비행의 특성으로 파악할 수 있다. 오프라인

공간에서는 사법 절차, 치료, 금전적 배상 등을 통해 피해에 대한 조치가 가능한 경우들이 존재하지만, 온라인 공간에서 유포된 이미지와 게시물은 사실상 완전한 삭제가 불가능하여 온라인 비행의 피해자들이 겪는 고통은 영구적으로 남을 수 있다(김신아, 방은혜, 한윤선, 2017).

### 이론적 배경

청소년의 비행을 설명하는 대표적인 이론으로 사회학습이론, 일반긴장이론, 사회통제이론, 일상활동, 낙인이론을 꼽을 수 있다. 본 연구는 해당 이론들을 통해 온·오프라인 상의 비행을 예측하는 다양한 요인들을 설명하고자 하였다.

#### 사회학습이론

사회학습이론(social learning theory)의 핵심은 청소년의 비행이 주변 사람들과의 상호작용을 통해 비행 행동을 학습한 결과로써 일어난다는 것이다(Akers & Jennings, 2015). 이론에서는 차별적 접촉(differential association), 정의(definition), 차별적 강화(differential reinforcement), 모방(imitation)이라는 네 가지 개념을 통해 청소년이 비행을 학습하고 행동으로 옮기는 메커니즘을 설명한다. 차별적 접촉은 청소년들이 오프라인 공간에서 만나는 사람들뿐만 아니라, 온라인 공간에서 접촉하는 대상으로부터 영향을 받는 것을 뜻한다(Warr, 2002). 예컨대, 청소년이 온라인 혹은 오프라인 상에서 비행에 우호적인 태도를 지닌 타인과 어울리거나 법률을 어기는 것에 거리낌이 없는 대상과 주로 접촉한다면 청소년 또한 이와 같은 행동에 가담할 위험이 증가하는 것이다(Akers

& Jennings, 2015). 정의는 범죄, 비행, 그리고 규범적인 행동에 대해 청소년이 부여하는 의미와 지니는 태도를 의미한다. 이론에 따르면 일탈 행위를 적절한 것으로 여기거나 행위를 정당화할 수 있다고 믿는 경우, 청소년이 비행을 저지를 가능성은 커진다. 또한, 비행에 대한 청소년의 우호적 태도는 비행에 가담할 위험을 높인다. 차별적 강화는 특정 행동에 뒤따르는 상벌에 대한 청소년의 인식, 과거 경험, 기대 간의 균형과 관련이 있다. 예를 들어, 비행을 통해 보상을 얻는 경험이 빈번하다면 이와 같은 행동이 강화되어 청소년이 비행을 지속할 가능성은 높아질 것이다(Akers & Jennings, 2015). 마지막 개념인 모방은 대리적(vicarious) 학습과도 관련이 있는데, 이론에 따르면 청소년의 행동은 타인의 행동과 그에 따르는 결과를 직접적·간접적으로 관찰함으로써 형성된다(황성현, 이강훈, 2013). 즉, 청소년은 타인이 비행을 저지르는 모습과 이후 결과를 관찰한 후에 유사한 행동을 할 수 있다. 선행연구에 따르면 비행에 가담하는 친구가 많을수록 청소년은 친구의 행동을 모방해 비행에 가담할 가능성이 증가하였다(Walters, 2018). 형제자매 또한 청소년이 모방할 수 있는 주된 대상으로(Farrington, 2010) 형제자매의 수는 청소년의 비행위험과 정적인 관계를 나타내었다(Farrington, 2010; Kierkus & Hewitt, 2009). 이와 관련해 선행연구들은 형제자매의 수가 많을수록 각각의 자녀에게 주어지는 부모 감독의 수준이 저하되거나(Farrington, 2010), 가족 구성원들의 밀집으로 갈등 상황이 상대적으로 자주 발생하여 청소년의 비행 가능성을 높인다고 보았다(Brownfield & Sorenson, 1994).

### 일반긴장이론

일반긴장이론(general strain theory)에 따르면 청소년이 비행을 저지르는 이유는 긴장 요인(strain)으로 인해 유발된 부정적인 감정을 해소하기 위해서이다(Agnew, 1992). 해당 이론은 개인의 병리적(pathological) 특성을 통해 범죄와 비행의 발생을 설명한 이론들을 반박하면서 등장하였는데, 누구나 스트레스를 유발하는 상황에 놓이는 경우 비행에 가담할 수 있다고 하였다(Agnew & Brezina, 2012). 청소년은 일상 속에서 다양한 긴장 요인들을 마주한다. 일반긴장이론에서 제시하는 긴장 요인은 목표 달성의 실패(the failure to achieve positively valued goal), 긍정적 자극의 제거(the removal of positively valued stimuli), 부정적 자극의 제시(the presentation of negative stimuli)의 세 가지이다. 먼저 목표 달성의 실패는 다시 세 가지 하위 항목으로 세분화할 수 있는데, 열망과 기대 간의 격차, 기대와 실제 성취 간의 격차, 공정한 결과와 실제 성취 간의 격차가 이에 해당한다(Agnew, 1992). 해당 긴장 요인에서 의미하는 목표는 돈, 또래 집단 내 지위, 성인들로부터의 자율성, 흥분감 등 다양하게 정의될 수 있다(Agnew & Brezina, 2012). 긍정적 자극의 제거는 청소년이 소중히 여기는 요소가 상실되는 것으로, 부모의 사망, 친구와의 이별 등이 대표적이다(황성현, 이강훈, 2013). 부정적 자극의 제시는 부모의 부적절한 양육 태도, 학대 및 방임, 또래와의 갈등, 범죄 피해 등 청소년에게 부정적 사건이 발생하는 것을 의미한다(Agnew & Brezina, 2012). 이와 같은 긴장 요인들은 청소년에게 분노, 우울, 불안과 같은 부정적인 감정들을 유발하고 청소년들은 부정적인 감정에서 벗어나고자 비행을 저지르게 된다는 것이 일반긴장이론의 설명이다

(Agnew, 1992). 선행연구들은 본 연구에서 긴장 요인으로 파악한 변인들이 청소년의 비행과 관련이 있음을 확인하였다. Marcal(2021)에 따르면 부모의 심리적 어려움은 청소년에게 부정적 정서를 불러일으키거나, 부정적 양육 태도를 통해 긴장을 유발하는 것으로 나타났다. 이혼 등으로 인한 가족구성의 변화 또한 청소년에게 심리적 어려움을 유발함으로써 비행 위험을 높일 가능성이 존재하였다(Alboukordi, Nazari, Nouri, & Sangdeh, 2012). 마찬가지로, 잦은 전학 경험은 청소년에게 스트레스를 유발할 수 있으며(Gasper, DeLuca, & Estacion, 2012) 이는 비행과 관련성을 나타내었다(Dudovitz et al., 2018).

### 사회통제이론

앞서 설명한 두 이론이 “왜 청소년은 비행에 가담하는가?”라는 질문에 대한 대답이었다면, 사회통제이론(social control theory)은 “왜 청소년은 비행에 가담하지 않는가?”에 초점을 둔다. 사회통제이론에 따르면 청소년들이 비행을 저지르는 이유는 단순하다. 목표를 달성하기 위해 규범적·합법적인 수단을 이용하는 것보다 비행을 저지르는 것이 훨씬 편리하기 때문이다(Hirschi, 1969). 이론에서는 비행의 동기가 기본적으로 모든 청소년에게 내재한다고 여기며 비행을 인간 본성에 부합하는 자연스러운 행동으로 파악한다(Hirschi, 1969). 하지만, 대부분의 청소년들은 비행을 저지르기보다 규범적인 행동을 택하는데, 사회통제이론은 이를 사회적 유대(social bond)를 통해 설명한다. 사회적 유대는 애착(attachment), 관여(commitment), 참여(involvement), 그리고 신념(belief)이라는 네 가지 요소를 포함한다. 먼저, 애착이란 청소년이 가족, 교사와 같은 중요한 타인과 맺는 정

서적인 유대 관계를 의미한다. 타인과 강한 유대 관계를 맺은 청소년들은 비행 행동이 관계에 부정적인 영향을 미칠 것을 우려하여 비행에 가담하지 않을 가능성이 높다. 실증연구에서는 부모, 교사와 같이 주요한 성인과 우호적인 관계를 맺는 경우 청소년의 비행위험이 감소하는 것으로 나타났다(Sabatine, Lippold, & Kainz, 2017; Obsuth, Murray, Knoll, Ribeaud, & Eisner, 2021). 관여는 교육적, 직업적 목표 등에 대해 청소년이 전념한 정도로, 비행은 청소년들이 과거에 성취한 것들을 잃게 할 수 있으며, 이에 따라 관여는 비행을 억제하는 역할을 한다. 참여는 방과 후 활동, 봉사활동과 같이 비행 행동을 저지하는 관습적 활동에 대한 참여를 의미한다. 관습적인 활동에 참여하는 것은 청소년이 비행을 저지할 수 있는 시간을 감소시킴으로써 비행을 억제한다. 마지막으로 신념은 청소년이 법률과 사회 내 규범이 타당하다고 여기는 정도로서(Costello & Laub, 2020), 법률과 규범을 따라야 한다는 믿음이 약할수록 청소년이 비행을 저지할 가능성은 커진다(Hirschi, 1969). 이외에도 공동체 의식이 낮을수록 비행위험이 증가하였으며 (Valasik & Barton, 2018), 지역사회 안전에 대한 인식이 부정적일수록 비행에 가담할 위험이 증가하였다(Hoffmann, Thorpe, & Dufur, 2020).

### 일상활동이론

일상활동이론(routine activity theory)에 따르면 청소년의 비행이 발생하기 위해서는 특정한 외부적 조건들이 충족되어야 한다(Cohen & Felson, 1979). 이론에서 제시하는 외부적 조건에는 가능한 보호력의 부재(the lack of a capable guardian), 동기화된 가해자(a motivated offender), 적절한 대상(a suitable target)이 있다(Cohen &

Felson, 1979). 구체적으로 비행은 피해 대상을 보호하거나 가해자를 저지할 수 있는 보호력의 수준이 낮고, 잠재적 피해 대상의 가치가 높거나, 그에 대한 접근이 용이하며, 비행 동기를 가진 청소년이 존재하는 경우에 발생 위험이 증가한다(Akers, Sellers, & Jennings, 2017). 예컨대, 청소년을 관리·감독하는 사람이 없는 상황에서 값비싼 물건이나 취약한 피해자를 발견한 경우, 청소년이 비행에 가담할 위험성은 높아진다. 한편, 이론을 검증한 후속 연구들은 비행을 저지르는 청소년들의 일상에는 특정한 패턴이 나타나는 것을 확인하였다. 구체적으로 부모가 청소년을 관리·감독하는 시간이 짧은 경우 비행 위험이 증가하였으며(홍명기, 민수홍, 조은비, 2020; Miller, 2013), 친구와 보내는 시간 또는 비구조화된 활동에 참여하는 시간의 증가는 비행 발생 가능성을 높이는 것으로 나타났다(Hoeben & Weerman, 2016; Hoeben, Osgood, Siennick, & Weerman, 2021). 휴대전화 의존도, 휴대전화 이용빈도 등은 보호력이 결여된 환경 속 청소년의 활동을 간접적으로 보여주는 것으로서 실증연구에서 온라인 공간에서의 활동이 잦거나(Van Ouytsel, Ponnet, & Walrave, 2018) 스마트폰에 과의존하는 생활양식을 가진 청소년은 사이버 비행에 가담할 가능성이 증가하는 것을 확인할 수 있다(김선영, 한윤선, 2022). 반면 독서, 공부, 학교 숙제와 같은 구조화된 활동에 보내는 시간은 비행과 부정 관계가 있었다(Agustina & Felson, 2015). 종합해보면, 일상활동이론은 비행 발생에 있어서 상황적 맥락의 중요성을 강조하는 이론임을 알 수 있다.

### 낙인이론

낙인이론(labeling theory)에 따르면 청소년을

지속적인 비행으로 이끄는 것은 공식적 기관의 통제와 사회 구성원들의 비공식적인 반응이다(박현수, 박성훈, 정혜원, 2009). 청소년이 사소한 문제 행동이나 비행을 저지르는 경우 공식기관으로부터의 제재나 주변 사람들의 부정적인 반응이 발생하고, 이는 낙인이 되어 청소년의 부정적인 자아상(self-image) 형성에 영향을 미친다(이성식, 전신현, 2009). 이와 같은 부정적 자아상은 '자기충족적 예언'으로 작용하여 후속 비행의 발생 가능성을 높일 수 있다(박현수 등, 2009). 또한, 낙인을 부여받은 청소년은 또래 집단으로부터 거부되는 등 규범적인 사회 구성원들로부터 배제되거나, 비행 성향을 지닌 친구들과 어울리면서 후속 비행 위험이 증가하기도 한다(Bernburg, 2019). 좁은 의미에서 낙인은 공식 낙인을 지칭하는 것으로, 교내 징계, 사법 체계로부터의 처벌 등에 의해 형성된 부정적 자아상을 의미한다(이성식, 전신현, 2009). 초기 낙인 이론가들의 주요 관심사는 공식 낙인에 한정되어 있었으나, 공식 낙인의 효과에 대한 비일관적인 연구 결과가 제시되면서 비공식 낙인이 주목을 받기 시작했다(박현수 등, 2009). 비공식 낙인이란 부모, 교사, 또래 등 주변 사람들로부터의 반응을 청소년이 지각하고 반영하여 형성되는 부정적 자아 개념이다(박현수 등, 2009; Matsueda, 1992). 예컨대, 교사나 부모에 의해 문제아라는 평가를 받은 청소년은 자신을 문

제이라고 규정하게 됨으로써 비행을 저지르는 가능성이 증가하는 것이다(Matsueda, 1992). 종단자료를 사용해 낙인이론을 검증한 실증연구들은 비공식 낙인이 앞서 제시한 경로들을 통해 청소년의 비행을 유발하거나, 처음으로 저지른 비행 이후에 청소년의 비행을 지속시킬 수 있음을 확인하였다(이성식, 전신현, 2009; 장은비, 최현주, 하규영, 박은혜, 이상민, 2014; 박현수 등, 2009; Bernburg, 2019).

## 방 법

### 연구대상

본 연구에서는 한국청소년정책연구원에서 제공하는 한국아동·청소년패널조사 2010을 활용하였다. 해당 조사는 아동·청소년의 다양한 성장과 발달 양상을 다각적으로 파악하기 위한 조사로써 초등학교 1학년, 4학년과 중학교 1학년을 모집단으로 하여 다단층화집락표집을 통해 표본을 추출하였으며, 2010년 1차 조사를 시작으로 2016년에 조사를 완료하였다. 1차 조사를 완료한 원패널은 총 7,071명(초1 패널 2,342명, 초4 패널 2,378명, 중1 패널 2,351명)이며 전체 패널 데이터의 측정 시점별 표본 수와 원표본 유지율은 표 1과 같다.

표 1. 전체 패널 데이터의 시점별 표본 수와 원표본 유지율

	1차 (2010)	2차 (2011)	3차 (2012)	4차 (2013)	5차 (2014)	6차 (2015)	7차 (2016)	
전체 패널	표본수(명)	7,071	6,808	6,678	6,319	6,271	6,184	5,862
	유지율(%)	100.0	96.3	94.4	89.4	88.7	87.5	82.9

주. 한국아동·청소년패널조사 2010 1차~7차 조사자료



본 연구에서는 예측모형의 성능을 높이기 위해 최대한 많은 사례를 연구대상에 포함하는 것을 목적으로 초1, 초4, 중1 패널의 6차년도 데이터를 병합하였다. 초1 패널의 경우, 6차년도 조사부터 온·오프라인 비행 관련 문항이 포함되어 본 연구에서는 모든 패널의 6차 조사자료를 사용하였다. 그리고 분석에 사용되는 비행 관련 변수들의 결측치가 없는 케이스만을 추려내어 총 4,137명(남: 2,035명, 여: 2,102명)의 사례를 최종적으로 분석에 활용하였다.

#### 측정도구

##### 반응변수<sup>2)</sup>

본 연구에서 반응변수로 활용하는 변수는 온라인 비행과 오프라인 비행이다. 원자료에 보고된 온라인 비행과 오프라인 비행은 각각 6문항(온라인 상에서 폭력적인 언어 사용, 불법 소프트웨어 다운로드, 해킹 등)과 14문항(음주, 무단결석, 가출, 폭행 등)으로 이루어졌으며, 비행을 경험한 경우 1, 경험하지 않은 경우 0으로 측정되었다. 각 세부 문항의 경험 유무를 기준으로 한 번이라도 경험한 경우를 1, 그렇지 않은 경우를 0으로 코딩하여 온라인 비행과 오프라인 비행이라는 이분형 변수를 재구성하였다. 온·오프라인 비행의 행동 유무별 빈도표는 아래 표 2와 같으며, 전반적으로 여러 비행 행동 유형 중 하나도 경험하지 않은 경우가 경험한 경우보다 상대적으로 높은 빈도를 보였다.

2) 반응변수(response variable)는 머신러닝 분석에서 통용되는 용어이며 예측변수(predicator variable)에 반응하는 변수로 종속변수와 동일한 개념으로 이해할 수 있다.

표 2. 청소년의 온·오프라인 비행 경험 빈도표  
(N=4,137)

		오프라인 비행	
		비행행동 무	비행행동 유
온라인	비행행동 무	3,129명(75.6%)	535명(12.9%)
	비행행동 유	326명(7.9%)	147명(3.6%)

##### 예측변수

본 연구는 비행과 관련된 예측변수 총 129개를 분석에 포함하였다. 반응변수가 오프라인 비행인 모형의 경우 오프라인 비행 행동 경험 변수(오프라인 비행 유무 14문항)들을 예측변수에서 제외하였으며, 반응변수가 온라인 비행인 모형에서는 온라인 비행 행동 경험 변수(온라인 비행 유무 6문항)들을 제외하여 각각 115개 및 123개의 예측변수를 모형에 활용하였다. 아래 표 3에 다수의 예측변수를 포함하는 비행 관련 이론을 기준으로 예측변수들을 범주화하여 제시하였다.

본 연구는 최대한의 사례 수를 유지하면서, 분석 데이터의 정보를 최대한 활용하여 청소년의 온·오프라인 비행 행동 경험을 예측하는 것에 중점을 두어 예측변수 관련 전처리를 진행하였다. 우선, 초1, 초4, 중1 패널 6차년도 데이터를 병합하여 세 패널 데이터에 공통적으로 포함된 변수만을 추려내었다. 그리고 반응변수를 예측하는데 불필요한 변수(표본ID, 학교ID, 패널 구분 등), 비행 관련 이론에 부합하지 않는 것으로 판단되는 변수(키, 몸무게, 용돈 주는 사람 등), 결측값이 20% 이상인 변수(해외 교육/연수 기간, 팬덤활동 빈도 등)를 제외하였다.

예측변수들 중 리커트 척도로 구성된 변수들은 양적변수로 취급하였으며, 문항이 갖는

표 3. 비행 이론별 예측변수

이론	하위요인	변수 수	예측변수
사회학습 이론	차별적 접촉	14	오프라인 비행 경험한 친구 수
	차별적 강화	20	오프라인 비행 경험, 온라인 비행 경험
	모방	5	형제자매 유무, 형제자매 수
일반긴장 이론	목표달성 실패	7	가구 연간 소득, 가정의 경제수준 평가, 건강상태 평가, 모친 근로 여부, 부친 근로여부, 월평균 용돈, 전체 성적 만족도
	긍정적 자극 제거	3	가족구성, 부모구성, 연간 전학 경험
	부정적 자극 제시	10	다문화가정 여부, 보호자의 건강상태 평가, 보호자의 삶의 만족도, 오프라인 비행 피해
	정서문제	6	삶의 만족도, 정서 문제(주의집중, 공격성, 신체증상, 사회적위축, 우울)
사회통제 이론	애착	8	교사관계, 교우관계, 또래애착, 양육방식, 이성친구 유무
	관여	6	학교규칙, 학습습관, 학습활동
	참여	5	가족/단체 여행 연간 경험, 문화활동 연간 경험 횟수, 방과 후 교육 연간 경험, 체육시간 중 운동시간, 해외방문 연간 경험 유무
	신념	3	공동체의식, 다문화수용성, 지역사회 인식
일상활동 이론	보호력 부재	20	기상시간, 기타 공부시간, 독서시간, 취침시간, 친구들과 노는 시간, 컴퓨터 게임/오락 시간, 학교 숙제 시간, 학원 과외 숙제 시간, 학원 과외 시간, TV 비디오 시청시간
	동기화된 가해자	10	휴대전화 의존도, 휴대전화 이용빈도
	적절한 대상	4	컴퓨터 사용여부, 팬덤활동, 휴대전화 보유여부, 휴대전화 종류,
낙인이론	부정적 자아상	2	자아정체감, 자아존중감
기타	-	6	거주지역, 교급, 모친 최종학력, 부친 최종학력, 성별, 출생년도

주1. 각 변수에 대한 자세한 척도는 부록에 제시함.

주2. 분류가 반드시 상호배타적이지 않음. 즉, 특정 이론에 분류된 예측변수들이 다른 이론에도 적용 가능함.

개념의 수준이 높아짐에 따라(혹은 정도가 강해짐에 따라) 문항 점수도 커지도록 일부 문항을 역코딩하였다. 또한 동일한 개념을 묻는 여러 하위 문항이 있는 경우 평균을 계산해 단일한 변수를 생성하였다. 평균 점수로 생성된 변수들은 교우관계, 우울, 공동체의식 등

총 25개이며 각 변수에 해당하는 문항들에 대한 크론바흐 알파(Cronbach's  $\alpha$ ) 값의 범위는 .66~.90으로 내적 일관성 신뢰도가 전반적으로 양호한 것으로 나타났다. 그 외 독서시간(등교/비등교), 학교숙제시간(등교/비등교) 등의 시간을 묻는 문항의 경우, 모두 시 단위로 통

일하였다.

마지막으로 단위가 변수 간에 상이한 문제를 해결하여 해석의 용이성을 높이고, 머신러닝 알고리즘의 예측 성능을 향상시키고자 모든 양적변수는 표준화를 진행하였다(Géron, 2019).

## 절차

본 연구는 청소년의 온라인 및 오프라인 비행 행동을 예측하기 위해 랜덤포레스트(random forest) 머신러닝 기법을 활용하였다. 랜덤포레스트 기법은 여러 의사결정나무 모형을 종합(ensemble)하는 분류 예측 기법이다(박소영, 정혜원, 2020; 손윤희, 박현정, 박민호, 2020). 단일한 의사결정나무 모형은 과적합 문제로 인해 안정성이 떨어진다는 단점이 있는데, 이를 보완하기 위해 랜덤포레스트 기법은 부트스트래핑(bootstrapping)을 통한 표본 추출을 통해 여러 의사결정나무 모형을 생성하고, 노드 분할 기준이 되는 변수를 무작위로 선정하여 예측력과 안정성을 높이는 전략을 취했다(Breiman, 2001; 유진은, 2015). 해당 기법은 비모수적(non-parametric) 모형을 기반으로 하여 독립변수의 외생성 가정과 같은 주요한 가정에서 자유로우며, 많은 수의 예측변수를 모형에 투입하더라도 다중공선성 및 자유도 감소 문제의 우려가 적다는 장점을 갖는다(홍기혜, 2020; 최필선, 민인식, 2018). 또한 이상치(outlier)의 영향이 적고, 오히려 다수의 예측변수가 투입됐을 때 예측 성능이 우수한 안정적인 모형이 구축된다. 일반적으로 머신러닝은 분석 결과에 대한 해석이 용이하지 않다는 단점이 있으나, 랜덤포레스트 기법은 중요도 지수(Mean Decrease Gini, MDG)와 부분의존성 도

표(partial dependence plots)를 통해 예측변수와 반응변수의 관계성을 시각화할 수 있으므로(유진은, 2015; 손윤희 등, 2020), 결과에 대한 해석도 용이하다.

랜덤포레스트 분석을 적용하기 위한 단계는 다음과 같다. 측정도구 파트에서 기술한 바와 같이 전처리 과정을 통해 변수 변환 및 결측치 제거를 완료한 데이터를 7:3의 비율로 훈련데이터와 검증데이터로 분리한다. 이때 데이터 분포의 편향을 막기 위해 무작위로 데이터를 분리하되 반응변수의 0(비행경험 무)과 1(비행경험 유)의 비율을 고려하여 비슷하게 나뉘도록 한다(周志華, 2020). 예를 들어, 반응변수에 대하여 0으로 응답한 사례가 100명이고, 1로 응답한 사례가 200명이라면, 이를 분리할 때, 훈련데이터에는 0으로 응답한 사례 70명, 1로 응답한 사례 140명을 포함하고, 검증데이터에는 0과 1의 사례가 각각 30명, 60명이 포함되게끔 무작위로 분리한다. 훈련데이터는 교차검증을 포함하여 예측모형을 구축하기 위해 활용되고, 검증데이터는 훈련데이터를 통해 구축된 예측모형의 성능을 평가하기 위해 활용된다. 훈련데이터를 통한 교차검증은 5겹 교차검증(5-fold-cross-validation)을 3번 반복하였으며, 검증데이터를 통한 예측모형의 평가는 정확도(accuracy)와 ROC-AUC-점수(ROC-AUC-score)를 활용하였다. 정확도는 예측모형에서 정확히 예측된 사례의 개수를 전체 사례로 나눈 값이고, ROC-AUC-점수는 1-특이도(false positive rate)<sup>3)</sup>가 x축이고 민감도(true

3) 특이도(specificity)는 실제 비행을 하지 않는 청소년들 중 예측모형에서 비행을 하지 않는 청소년으로 정확히 예측한 경우의 비율이고, 민감도(sensitivity)는 실제 비행을 하는 청소년들 중 예측모형에서 비행을 하는 청소년으로 정확하게

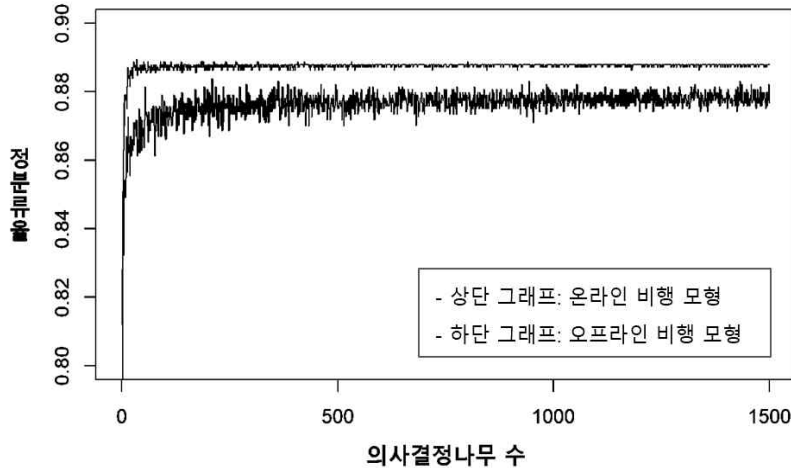


그림 1. 온·오프라인 비행 모형별 의사결정나무 수에 따른 정확도

positive rate)가  $y$ 축인 ROC-곡선(receiver operating characteristic-curve) 아래의 영역(area under the curve)을 의미한다(Géron, 2019). ROC-AUC-점수는 0.5~1.0의 범위를 가지며 0.7~0.8은 ‘보통’, 0.8~0.9는 ‘ 좋음’, 0.9~1.0은 ‘매우 좋음’으로 평가할 수 있다(Muller et al., 2005). 평가 과정에서 랜덤포레스트 모형의 적절성을 검증하기 위해 다른 지도학습 머신러닝 기법인 서포트벡터머신(support vector machine), 신경망모형(neural network model), 그래디언트 부스팅(gradient boosting)의 예측 성능과 비교하였다.<sup>4)</sup>

예측한 비율이다(서민구, 2014).

4) 서포트벡터 머신은 분류하고자 하는 집단 사이의 거리(margin)를 최대로 하는 경계면을 찾는 방식으로 모형이 구축된다. 신경망 모형은 뇌신경계를 모방한 방법으로서 뉴런에 해당하는 노드로 구성된 입력층, 은닉층, 출력층이 연결되어 분류하고자 하는 집단을 최대한 분류할 수 있도록 각 노드의 활성화 함수와 관련된 가중치를 조정하는 방식으로 모형이 구축된다. 그래디언트 부스팅은 다수의 기초학습기를 결합하는 앙상블 방식을 적용하지만 랜덤포레스트와는 달리 기초

다음으로 랜덤포레스트의 하이퍼파라미터 튜닝(hyperparameter tuning)과 관련하여, 기초학습기인 의사결정나무를 몇 개 활용할지와 의사결정나무에서 노드를 분리할 때 무작위로 선택되는 예측변수의 수를 결정하였다. 예측변수의 수는 일반적으로 전체변수 수의 제곱근을 권장하는데(Díaz-Uriarte & Alvarez de Andrés, 2006), 본 연구에서는 해당 숫자(11개)를 포함하여 그리드 서치(grid search)를 통해 1~30개의 변수의 수 중 최적의 예측 성능을 보이는 것을 선택하도록 설정하였다. 의사결정나무의 수의 경우, 아래 그림 1과 같이 1개에서 1,500개로 의사결정나무의 수가 증가함에 따라 정확도가 어떻게 변화하는지 사전 탐색을 하였다. 그 결과 1,500개의 의사결정나무는 본 연구의 분석 데이터에서 정확도를 일정 수준으

학습기가 순차적으로 생성되며, 손실함수의 최소화를 목적으로 이전 학습기의 결과를 통해 가중치가 조정되어 이후 학습기에 적용되는 방식으로 최종 모형이 구축된다(박소영, 정혜원, 2020; 周志華, 2020).

로 수렴하게 하는데 충분한 숫자로 판단하여, 해당 개수로 분석을 진행하였다.

전술한 과정을 통해 랜덤포레스트 분석을 진행하였고, 구축된 두 모형(온라인 비행 모형, 오프라인 비행 모형)에 대하여 중요도 지수를 기준으로 예측에 주요한 역할을 한 상위 10개의 변수를 탐색하였다. MDG는 여러 의사결정 나무가 각각 노드를 분리하는 과정에서 특정 변수를 기준으로 노드가 나뉘어졌을 때 그 이전보다 분류가 더 잘 수행된 정도(불순도의 감소 정도)를 평균한 값이다(최필선, 민인식, 2018). 마지막으로, 상위 10개의 주요 변수들과 반응변수의 관계성을 살펴보기 위해 부분 의존성 도표를 활용하여 시각화하였다. 부분 의존성은 특정 예측변수가 반응변수에 미치는

한계효과(marginal effects)를 말한다(홍기혜, 엄태호, 2021). 분석을 위한 전처리는 SPSS 26과 R 4.0.5를 사용하였고, 머신러닝 분석은 R의 caret 패키지(Kuhn, 2008)를 활용하였다.

## 결 과

### 예측 모형의 성능 평가

훈련데이터를 통해 구축된 예측모형을 평가하기 위해 검증데이터를 투입하여 정확도와 ROC-AUC-점수를 도출하였고, 그 결과는 표 4와 같다. 정확도의 경우 온라인 비행 모형은 랜덤포레스트의 결과와 다른 머신러닝 모형들

표 4. 머신러닝 기법별 평가지표

	랜덤포레스트		서포트벡터머신		신경망모형		그래디언트부스팅	
	온라인	오프라인	온라인	오프라인	온라인	오프라인	온라인	오프라인
정확도(accuracy)	.89	.88	.89	.86	.89	.86	.89	.87
ROC-AUC- 점수	.72	.87	.52	.83	.69	.83	.72	.87

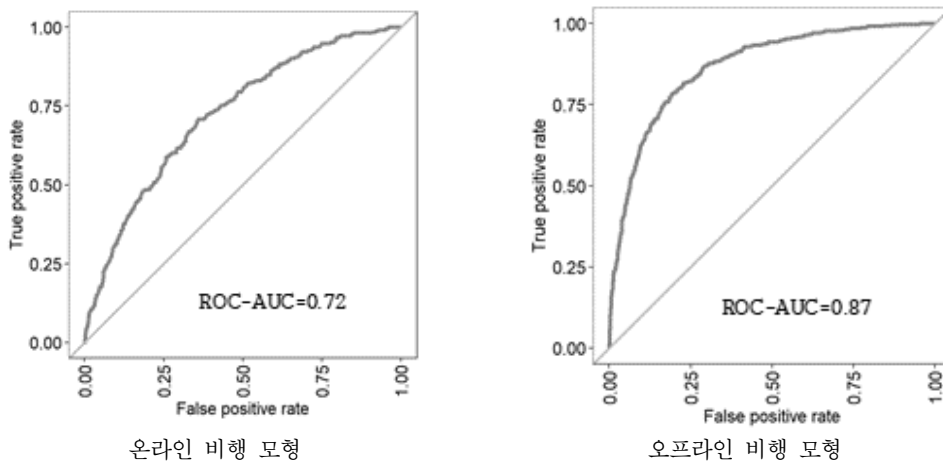


그림 2. 랜덤포레스트 모형별 ROC-곡선 및 ROC-AUC-점수

의 결과가 같은 수준을 보였고, 오프라인 비행 모형에서는 랜덤포레스트가 가장 높은 수준을 보였다. ROC-AUC-점수의 경우, 온라인 및 오프라인 비행 모형 모두에서 랜덤포레스트와 그래디언트부스팅이 가장 높은 수치를 나타냈다. 따라서 종합적으로 살펴보면, 랜덤포레스트의 예측 성능이 다른 머신러닝 기법에 못지않고 ROC-AUC-점수가 온라인 비행 모형에서는 0.7 이상, 오프라인 비행 모형에서는 0.8 이상으로 나타나 두 모형이 청소년의 온라인 및 오프라인 비행 행동을 예측하는 모형으로 기능할 수 있을 것이라 판단하였다.

#### 예측변수의 중요성 지수

온라인 비행 모형과 오프라인 비행 모형 각각에 대하여 중요도 지수를 기준으로 예측에 주요한 역할을 한 상위 10개의 변수를 그림 3과 그림 4에 나타내었다. 먼저 온라인 비행을 예측함에 있어 학교에 등교하지 않는 날의 컴퓨터 게임 오락시간이 가장 중요도가 높은 것으로 나타났다. 뒤이어 자아존중감이 높은 중

요도를 보였고, 다음으로 가구 연간 소득, 자아정체감, 신체증상, 기상시간(비등교), 성취가치, 휴대전화 의존도, 주의집중 문제, 공격성 순으로 높은 중요도를 나타냈다. 해당 결과를 통해 온라인 비행을 예측하는 주요 변인들이 전반적으로 일상활동이론, 일반긴장이론, 낙인 이론에서 자주 거론되는 청소년의 위험 행동을 설명하는 요인에 포함됨을 알 수 있다.

오프라인 비행 모형의 경우, 음주를 경험한 친구의 수가 가장 주요한 예측변수로 나타났고, 흡연을 경험한 친구의 수가 그 뒤를 이었는데, 해당 두 변수들은 중요도에 있어 나머지 변수들과 큰 격차를 보였다. 다음으로 학교규칙, 학습활동, 출생년도, 주의집중 문제, 성취가치, 기상시간(등교), 우울, 친구들과 노는 시간(비등교) 순으로 높은 중요도를 나타냈으며, 이들 변수는 비슷한 수준의 중요도를 보였다. 오프라인 비행을 예측하는 것으로 나타난 이러한 변인들은 일반긴장이론, 사회학습이론, 사회통제이론에서 청소년 비행의 중요한 요인으로 꼽히고 있다.

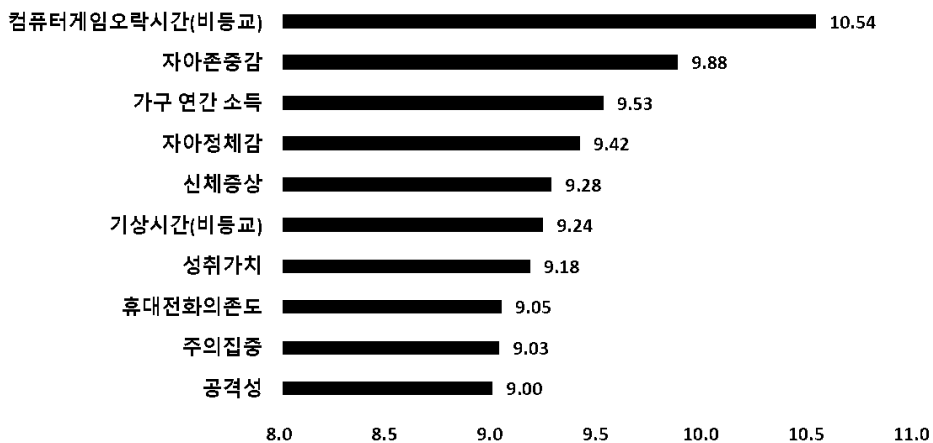


그림 3. 온라인 비행 모형의 주요변수(상위 10개)

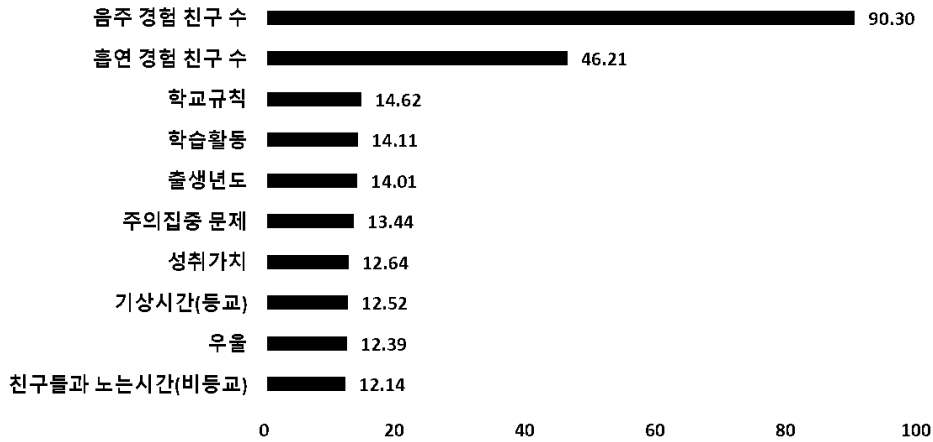


그림 4. 오프라인 비행 모형의 주요변수(상위 10개)

부분 의존성 도표

부분 의존성 도표를 통해 예측변수와 반응 변수 간의 관계를 시각화하였으며, 온라인 및 오프라인 비행 모형의 주요 예측변수 상위 10 개에 대한 부분 의존성 도표를 그림 5와 그림 6에 제시하였다. 양적 변수에 대해서는 표준

화를 진행하였으므로, x축에서 0에 해당하는 값은 해당 변수의 평균을 의미한다. 결과를 좀 더 구체적으로 살펴보면, 온라인 비행 모형의 경우 컴퓨터 게임/오락시간(비등교), 가구연간소득, 주의집중문제, 신체증상, 휴대전화 의존도, 공격성의 수준이 평균보다 높을수록 온라인 비행을 경험할 확률이 높은 것으로 나

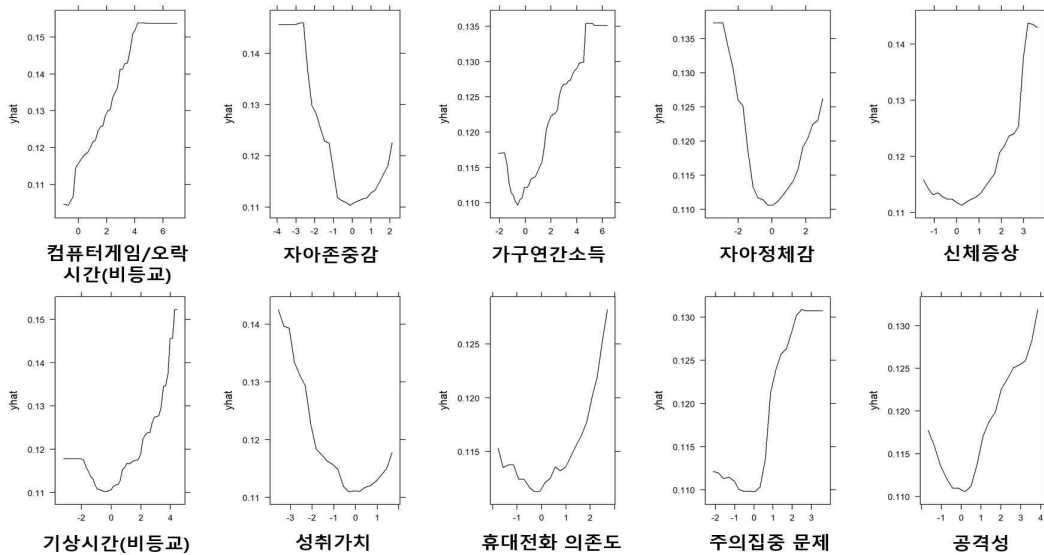


그림 5. 주요 변수의 부분의존성 도표(온라인 비행 모형)

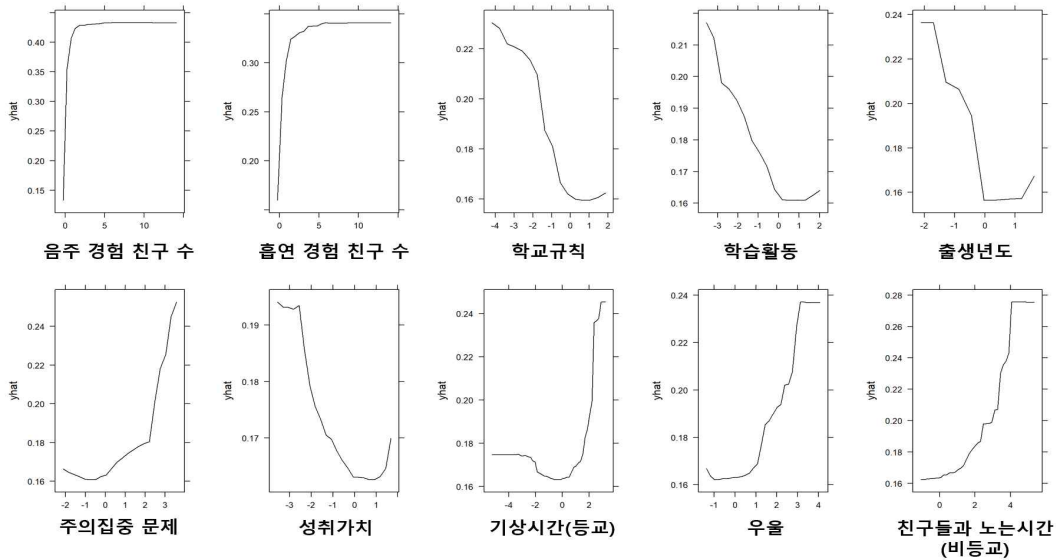


그림 6. 주요 변수의 부분의존성 도표(오프라인 비행 모형)

타났다. 기상시간(비등교)의 경우 평균보다 더 늦게 기상할수록 온라인 비행을 경험할 확률이 높았으며, 자아존중감과 자아정체감은 평균을 기점으로 낮은 수준을 보이거나 높은 수준을 보일수록 온라인 비행을 경험할 확률이 높아졌고, 상대적으로 평균보다 낮은 수준을 보이는 경우가 비행 경험 확률이 더 높았다.

오프라인 비행 모형의 경우, 음주 경험 친구 수, 흡연 경험 친구 수, 주의집중 문제, 친구들과 노는 시간(비등교), 우울 수준이 높을수록 오프라인 비행을 경험할 확률이 높아지는 것으로 나타났다. 기상시간(등교)의 경우 평균보다 늦게 일어날 경우 오프라인 비행 경험 확률이 높아졌으며, 학교규칙, 학습활동, 성취가치 수준이 평균보다 낮을수록 오프라인 비행 경험 확률이 높은 것으로 나타났다. 마지막으로 출생년도의 경우 연령이 높을수록 오프라인 비행 경험 확률이 높은 것으로 나타났다.

## 논 의

정보통신매체의 사용이 증가함에 따라 청소년 활동영역이 온라인과 오프라인 영역을 넘나드는 가운데, 본 연구는 랜덤포레스트 머신러닝 기법을 활용해 청소년 온라인 및 오프라인 비행을 예측하는 모형을 구축하였다. 나아가 두 가지 모형에서 제시하는 주요 변인들의 예측력을 확인함으로써 정보화 시대를 살아가는 청소년들의 행동을 과거에 정립된 비행 이론으로 설명 가능한지 탐색하였다. 분석 결과는 다음과 같이 종합할 수 있다. 온라인 비행의 경우, 일상활동이론에 기반한 통신기기 활용 관련 변수와 일반긴장이론, 낙인이론과 관련된 개인의 심리·정서적 변수들이 주요 예측요인으로 나타났다. 반면, 오프라인 비행에서는 일반긴장이론으로 설명할 수 있는 정서적 요인 등이 주요한 변인으로 확인되었으며, 또래의 지위비행, 청소년 학업 및 학교생활과



관련된 요인들이 예측변인으로 나타나 사회학 습이론과 사회통제이론의 설명력이 부각되었다. 각 비행 예측요인에 대한 구체적인 논의는 다음과 같다.

먼저, 일반긴장이론은 온라인 비행과 오프라인 비행 모두를 예측하는 중요한 이론적 기반임을 확인할 수 있었다. 온라인 비행의 경우, 정서문제로 대변되는 ‘주의집중’, ‘공격성’, ‘신체증상’이 주요 예측변인으로 도출되었으며, 오프라인 비행을 예측하는 주요 변수로 ‘주의집중’과 ‘우울’이 꼽혔다. 비행과 정서문제의 관계를 살펴본 선행연구들은 온라인과 오프라인 비행 모두 일반긴장이론을 기반으로 하는 경우가 많다(이석영 등, 2015; 이웅택, 이은경, 2021; 유석봉, 이완희, 2018; 최진실, 장현석, 2017). 본 연구결과도 이와 마찬가지로, 청소년이 일상에서 경험하는 여러 긴장 요인들로 인해 주의집중, 공격성, 신체증상, 우울 등의 정서문제를 경험하게 되면 이러한 부정적 정서를 완화하기 위한 대처수단으로 온라인과 오프라인 비행에 가담하는 것으로 설명할 수 있을 것이다.

다음으로 온라인 비행을 예측하는 모형을 먼저 살펴보면, 일상생활이론과 밀접하게 관련이 있는 ‘등교하지 않는 날의 컴퓨터 오락 시간’이 가장 높은 정적 영향력을 나타냈다. 컴퓨터 오락시간은 다른 주요 요인인 ‘휴대전화 의존도’와 함께 일상활동이론과 밀접하게 연결된다. 즉, 청소년의 일상생활에서 컴퓨터나 휴대전화 등의 전자기기를 활용한 비구조화된 온라인 활동이 차지하는 비중이 온라인 비행 경험을 설명하는 주요 요인으로 작용하고 있음을 알 수 있다. 이는 선행 연구에서 청소년의 인터넷 사용시간(이성식, 2009; 정혜원, 2009)과 청소년의 스마트폰 중독(이준형,

최용렬, 2020; 조춘범, 2017)이 사이버 비행에 정적 영향력을 갖는다고 나타난 것과 동일한 결과이다.

다음으로 ‘자아존중감’은 ‘자아정체감’과 함께 온라인 비행을 예측하는 주요 요인으로 꼽혔으며, 이는 낙인이론과 연결된다. 낙인이론은 외부에서 공식 혹은 비공식적으로 주어지는 ‘비행 청소년’이라는 낙인이 부정적 자아개념을 형성하고 이에 따라 추가적인 비행이 일어날 수 있다고 제시한다(김소희, 황의갑, 2015). 본 연구의 결과에서 자아존중감과 자아정체성 모두 평균을 중심으로 높거나 낮을 때 온라인 비행 경험이 증가하는 경향성을 나타냈다. 부정적 자아개념이 형성되어 있는 경우, 비행을 문제해결을 위한 수단으로 택하거나, 비행으로 얻는 이익을 손해보다 크다고 여길 가능성이 있다. 한편 자아개념이 지나치게 긍정적인 청소년의 비행 위험이 증가하는 것은 나르시시즘과 관련하여 이해할 수 있을 것이다. 나르시시즘은 과장되고 불안정한 자신을 우호적으로 평가하는 것을 의미하며, 자기중심적 사고에 초점을 둔다(박정민, 1998). 특히 나르시시즘 관련 선행연구에서는 지나치게 우호적으로 편향된 자기에적 성향이 공격성 및 폭력성과 관련되며 비행에도 정적인 영향을 주는 것으로 보고하고 있다(Baumeister, Smart, & Boden, 1996; Barry, Grafeman, Adler, & Pickard, 2007).

청소년의 오프라인 비행에 영향을 미치는 변수는 온라인 비행을 예측하는 주요 변수들과 차이를 나타내었다. 특히 전통적으로 청소년 비행을 설명하는 3대 이론으로 꼽히는 일반긴장이론, 사회학습이론, 사회통제이론(Agnew & Brezina, 2012)과 관련된 변수들의 예측 중도도가 높게 나타났다. 오프라인 비행을 예측하

는 변수 중 중요도가 가장 높은 두 변수는 ‘음주 경험 친구 수’와 ‘흡연 경험 친구 수’로, 이는 사회학습이론에서 제시하는 내용과 부합한다. 사회학습이론에 따르면 청소년은 비행 친구와의 차별적 접촉을 통해 비행 행동을 합리화하는 가치관을 습득하고, 나아가 직접 비행에 가담하게 된다(황성현, 2015; Akers et al., 2017). 이처럼 오프라인 공간에서 비행 친구들과 잦은 교류는 청소년에게 비행을 정당화할 수 있게 하고, 나아가 청소년이 직접 비행을 저지르는 결과로 이어질 수 있다. 또한, 비행 친구들이 청소년의 비행에 긍정적인 피드백을 제공하는 경우에는 청소년의 비행이 더욱 강화될 것이다.

사회통제이론에 기반한 ‘학교규칙’, ‘학습활동’, ‘성취가치’ 변수도 오프라인 비행 예측에 높은 중요도를 보였다. 이들 변수는 특히 사회통제이론에서 유대와 관련하여 제시한 네 가지 개념 중 청소년이 통상적인 목적을 달성하기 위해 투자한 노력과 시간을 가리키는 관여의 개념과 관련된 것이다(민수홍, 2018; 홍명기 등, 2020). 본 연구 결과와 관련하여 학교규칙을 준수하고, 학습활동에 적극적으로 참여하며, 성취에 대해 큰 가치를 부여하는 것은 관여의 정도가 높음을 의미한다. 관여가 높은 청소년은 비행에 가담하는 것이 자신이 투자한 시간과 노력에 큰 손실을 일으킨다는 것을 인지하여 비행에 가담할 위험성이 낮다. 한편, 학교규칙을 준수하지 않고 학습활동에 대한 참여도와 성취가치가 낮은 청소년은 순응에의 이해관계(stake in conformity) 또한 낮아 비행을 저지를 가능성이 높을 것이다.

본 연구의 결과 해석에 있어서 고려해야 하는 제한점은 다음과 같다. 첫째, 각 이론을 대표하는 변인들의 수가 동일하지 않기 때문에,

상대적으로 더 많은 변수를 포함하는 비행 이론이 더 큰 설명력을 갖는다는 해석의 오류를 범할 가능성이 크다. 따라서 추후 연구에는 각 이론에 연관되는 변수의 수를 고르게 배분한 분석이 필요할 것이다. 둘째, 본 연구에서 예측 변인마다 하나의 이론을 적용하였으나, 변인들은 다른 이론을 통해서도 설명될 수 있다. 따라서, 결과 해석 시 특정 이론과 관련지어 살펴본 변인이 다른 이론과 배타적이지 않음을 고려해야 한다. 셋째, 조사자료의 한계로 인해 본 연구결과가 실제 비행 고위험군 청소년의 응답을 대표하지 않았을 가능성이 존재한다. 본 연구에서는 1차년도에 비해 조사대상자들이 일부 누락된 6차년도 패널 자료를 활용하였는데 패널 자료의 특성상 위기 요인이 상대적으로 많은 청소년일수록 조사에서 누락되는 경향이 있기 때문이다. 그리고 조사자료의 초기 표본 설계과정에서 조사대상을 학교에 재학 중인 일반 청소년으로 하였으므로, 학교밖 청소년 등 국내 청소년들 모두의 온라인 및 오프라인 비행경험을 데이터가 온전히 반영하지 못했을 가능성이 있다. 넷째, 본 연구 결과에서 제공하는 반응변수와 예측 변수의 관계는 횡단자료에서 도출되어 인과관계가 아님에 유의해야 한다. 마지막으로, 주요 예측 변인을 도출할 때 사용된 기준은 중요도 지수의 상대적 크기이므로 해당 지수의 크기 자체에 대한 통계적 검증은 포함되지 않았음을 유의할 필요가 있다. 따라서, 추후 연구에서는 종단적 자료를 활용한 분석과 더불어 도출된 변인에 대한 통계적 유의성을 검증하는 분석을 추가해야 할 것이다.

전술한 제한점에도 불구하고, 본 연구는 랜덤포레스트 머신러닝 기법을 활용함으로써 100여 개가 넘는 예측변수를 통해 청소년의

온·오프라인 비행을 예측하는 모형을 도출하였다는 데 의의가 있다. 즉, 청소년 비행이라는 사회현상을 예측함에 있어 관련된 많은 변수들의 상호작용과 비선형성을 고려하는 고차원의 데이터를 활용함으로써 예측 성능 및 일반화 가능성을 향상시킨 결과를 도출한 것이다(홍기혜, 2020). 지금까지 많은 선행연구에서는 주로 청소년 비행을 예측함에 있어서 다른 변인들과의 다차원적 상호작용을 고려하지 않고 일부 변인들의 설명력을 파악함으로써 비행 위험 요인에 대한 부분적인 통찰을 제공하였다. 이와 달리, 본 연구는 랜덤포레스트 머신러닝 기법을 통해 청소년을 내포하고 있는 다양한 생태체계 요인들과 이들 간의 복잡한 상호작용을 종합적으로 고려함으로써 예방 및 개입 프로그램을 위한 포괄적 근거를 제시할 수 있었다. 나아가, 본 연구는 각 비행경험에 공통적으로 영향을 주는 요인과 차별적으로 영향을 주는 요인을 탐색함으로써 현 청소년 비행 현상에 대한 기존 비행이론의 적용 가능성을 함께 검증했다는 의의를 지닌다. 본 연구의 결과는 온라인 비행의 경우 청소년의 생활 패턴 및 심리·정서적 상태에 따라 비행 발생 가능성이 달라지는 반면, 오프라인 비행은 비행 또래 집단으로부터 큰 영향을 받으며, 청소년이 부여하는 관습적인 가치와 활동에 대한 중요성이 낮을수록 비행 발생 위험이 증가함을 제시하였다. 이처럼 온라인 및 오프라인 비행 가해 가능성이 상대적으로 높은 청소년의 특징에 대한 탐색은 예방 및 개입 프로그램을 개발하고, 학교 현장에서 이러한 청소년들을 선제적으로 발굴하기 위한 가이드라인을 개발하는 데에 참고자료로 활용될 수 있을 것이다.

## 참고문헌

- 과학기술정보통신부 (2021). 2021년 스마트폰 과의존 실태조사.
- 경찰청 (2019). 2018 경찰통계연보.
- 경찰청 (2020). 2019 경찰통계연보.
- 경찰청 (2021). 2020 경찰통계연보.
- 김선영, 한윤선 (2022). 부모의 스마트폰 의존이 청소년 사이버 비행에 미치는 영향: 부정적 양육 태도와 청소년 스마트폰 의존의 매개 효과. *청소년복지연구*, 24(2), 25-46.
- 김신아, 방은혜, 한윤선 (2017). 청소년 사이버 불링 가해행동 예측요인 탐색을 위한 국내연구 메타분석. *한국콘텐츠학회논문지*, 17(5), 18-33.
- 김소희, 황의갑 (2015). 학교 내 공식처벌이 청소년 비행에 미치는 영향: 자아낙인의 매개효과를 중심으로. *한국범죄학*, 9(3), 103-129.
- 김준범, 장대연, 최장원, 서정열 (2018). 청소년의 중비행에 관한 연구: 촉발과 반복에 영향을 미치는 정신건강요인을 중심으로. *청소년문화포럼*, 55, 27-51.
- 김준호, 노성호, 이성식, 광대경, 박정선, 이동원, ... 최수형 (2013). *청소년비행론*. 서울: 청목출판사.
- 김판준 (2019). 랜덤포레스트를 이용한 국내 학술지 논문의 자동분류에 관한 연구. *정보관리학회지*, 36(2), 57-77.
- 민수홍 (2018). 사회통제이론의 관점에서 본 수형자의 가족관계: 가족접견, 서신교환, 구금 후 가족관계 변화. *교정연구*, 28(2), 131-157.
- 박선숙 (2018). 청소년 개인 특성별 사이버비

- 행 위협군과 안전군에 관한 연구: 의사결정트리 분석을 이용하여. *사회과학연구*, 34(2), 53-79.
- 박소영, 정혜원 (2020). 중학생의 진로 결정 예측변수 탐색: 머신러닝 기법 적용. *아시아교육연구*, 21(3), 727-753.
- 박영신, 김의철 (2010). 친구가 청소년의 일탈 행동에 미치는 영향: 자기효능감, 부모자녀 갈등 및 부모의 통제를 중심으로. *한국심리학회지: 문화 및 사회문제*, 16(3), 385-422.
- 박정민 (1998). 비행 청소년의 나르시시즘, 지배성, 공격성 간의 관계. *이화여자대학교 대학원 석사학위 청구논문*.
- 박지수, 김하영, 유진경, 한윤선 (2017). 잠재계층분석(LCA)을 이용한 청소년-또래 비행의 유형과 특성. *아동학회지*, 38(2), 165-176.
- 박현수, 박성훈, 정혜원 (2009). 청소년비행에 있어 낙인의 효과에 대한 경험적 연구: 비공식 낙인을 중심으로. *한국청소년연구*, 20(1), 227-251.
- 방송통신위원회 (2021). 2021년 사이버폭력 실태조사.
- 법무연수원 (2022). 2021 범죄백서.
- 서경현 (2002). 청소년들의 데이트 폭력 가해 행동에 대한 사회학습적 변인들과 분노의 역할. *한국심리학회지: 문화 및 사회문제*, 8(2), 1-15.
- 서민구 (2014). R을 이용한 데이터 처리 & 분석 실무. 서울: 길벗.
- 손윤희, 박현정, 박민호 (2020). 랜덤 포레스트를 활용한 읽기소양 수준에 따른 집단 결정요인 분석: PISA 2018 자료를 중심으로. *아시아교육연구*, 21(1), 191-215.
- 양준영, 김지원, 김수영, 홍세희 (2019). 잠재전이분석을 적용한 청소년 사이버비행의 잠재계층 분류 및 전이 영향요인 검증. *청소년학연구*, 26(2), 101-130.
- 유석봉, 이완희 (2018). 일반긴장이론의 긴장요인과 사이버비행. *한국공안행정학회보*, 27(4), 279-306.
- 유진은 (2015). 랜덤 포레스트. *교육평가연구*, 28(2), 427-448.
- 여성가족부 (2020). 2020년 청소년 매체이용 및 유해환경 실태조사.
- 이석영, 이택호, 한윤선 (2015). 청소년의 폭력 비행 피해가 사이버 비행에 미치는 영향: 공격성, 우울, 사회적 위축의 매개효과. *한국아동복지학*, (50), 27-53.
- 이성식 (2008). 청소년 사이버범죄 유형별 원인 및 통제요소에 관한 통합적 연구. 서울: 형사정책연구원.
- 이성식 (2009). 초등학교생들의 사이버비행에 있어서 일상긴장, 낮은 자기통제력, 인터넷 사용시간의 영향작용: 세 이론의 검증을 중심으로. *형사정책연구*, 95-115.
- 이성식, 전신현 (2009). 학업성적과 청소년비행: 청소년패널자료를 통한 비행이론들의 검증. *한국청소년연구*, 20(2), 91-111.
- 이은정, 조희숙, 송영수 (2020). 랜덤 포레스트를 활용한 대졸 신입사원 조기이직 예측 결정요인 탐색. *기업교육과 인재연구*, 22(1), 163-193.
- 이용택, 이은경 (2021). 일반긴장이론에 근거한 청소년기 부정적인 사회적 관계와 사이버비행의 관계에서 공격성과 우울의 매개효과. *미래청소년학회지*, 18(3), 51-75.
- 이종원, 이순래, 정윤미 (2016). *한국아동·청소년패널조사 VII: 기초분석보고서 I-청소년비행의 실태와 추이 분석: 현실비행과 사이*

- 버비행. 세종: 한국청소년정책연구원.
- 이준형, 최응렬 (2020). 청소년의 스마트폰 중독이 사이버비행에 미치는 영향-부모·교사·또래애착의 병렬적 다중매개효과 검증. *한국중독범죄학회보*, 10(3), 23-40.
- 임미령, 김지훈, 김미희, 이현, 정애경 (2015). 경기지역 중학생 비행행동 잠재집단분석. *청소년학연구*, 22(7), 421-443.
- 장은비, 최현주, 하규영, 박은혜, 이상민 (2014). 성별에 따른 부모-자녀 관계, 비공식 낙인, 비행의 관계. *한국심리학회지: 학교*, 11(1), 185-205.
- 정혜원 (2009). 사이버비행 지속에 미치는 영향: 사회유대이론, 자기통제이론, 비행기회이론을 중심으로. *정보화정책*, 16(4), 97-112.
- 조춘범 (2017). 청소년의 스마트폰 중독이 오프라인 비행에 미치는 영향: 온라인 비행의 매개효과 검증. *한국사회복지조사연구*, 55, 317-345.
- 최진실, 장현석 (2017). 또래관계에서의 소외가 사이버비행에 미치는 영향: 일반긴장이론과 비행기회이론을 통하여. *한국범죄학*, 11(3), 3-17.
- 최필선, 민인식 (2018). 머신러닝 기법을 이용한 대졸자 취업예측 모형. *직업능력개발연구*, 21(1), 31-54.
- 홍기혜 (2020). 랜덤포레스트 머신러닝 알고리즘 기반 남·여 청소년의 자살생각 예측 및 분석. *한국사회복지학*, 72(3), 157-180.
- 홍기혜, 엄태호 (2021). 머신러닝 기반 복지재원 부담 태도 예측 및 분석: 세대별 비교를 중심으로. *사회과학논집*, 52(2), 91-114.
- 홍명기, 민수홍, 조은비 (2020). 다문화 청소년의 비행 원인에 대한 연구: 사회통제이론과 일상활동이론을 중심으로. *형사정책*, 32(3), 223-259.
- 홍정윤, 이수정 (2020). 잠재프로파일 분석을 통한 사이버 비행청소년 유형 분류 및 영향요인 탐색. *한국범죄학*, 14(2), 63-81.
- 황성현, 이강훈 (2013). 청소년 비행의 원인에 관한 사회학습, 사회유대, 일반긴장이론적 접근. *한국청소년연구*, 24(3), 127-145.
- 황성현 (2010). 청소년 지위비행에 관한 비행이론 비교연구. *청소년학연구*, 17(5), 49-68.
- 황성현 (2015). 청소년 비행이론의 상대적인 영향력 검증: 아동·청소년 패널자료를 중심으로. *한국경호경비학회지*, (44), 225-250.
- 周志華 (2020). 단단한 머신러닝. 김태현 (번역). 경기: 제이펍. (원전은 2016년에 출판)
- Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology*, 30(1), 47-88.
- Agnew, R., & Brezina, T. (2012). *Juvenile delinquency: Causes and control*. New York: Oxford University Press.
- Agustina, J. R., & Felson, M. (2015). Routine activities, delinquency, and youth convergences. In Alex R. Piquero (Ed), *The handbook of criminological theory* (pp. 137-155). UK: John Wiley & Sons.
- Akers, R. L., & Jennings, W. G. (2015). Social learning theory. In Alex R. Piquero (Ed.), *The handbook of criminological theory* (pp. 230-240). UK: John Wiley & Sons.
- Akers, R. L., Sellers, C. S., & Jennings, W. G. (2017). *Criminological Theories: Introduction, Evaluation, and Application 7th Edition*, UK: Oxford University Press.
- Alboukordi, S., Nazari, A. M., Nouri, R., &

- Sangdeh, J. K. (2012). Predictive factors for juvenile delinquency: the role of family structure, parental monitoring and delinquent peers. *International Journal of Criminology and Sociological Theory*, 5(1), 770-777.
- Barry, C. T., Grafeman, S. J., Adler, K. K., & Pickard, J. D. (2007). The relations among narcissism, self-esteem, and delinquency in a sample of at-risk adolescents. *Journal of Adolescence*, 30(6), 933-942.
- Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, 103(1), 5-33.
- Bernburg, J. G. (2019). Labeling theory. In J. G. Bernburg (Eds.), *Handbook on crime and deviance* (pp. 179-196). New York: Springer.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Brownfield, D., & Sorenson, A. M. (1994). Sibship size and sibling delinquency. *Deviant Behavior*, 15(1), 45-61.
- Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4), 588-608.
- Costello, B. J., & Laub, J. H. (2020). Social control theory: The legacy of Travis Hirschi's causes of delinquency. *Annual Review of Criminology*, 3(1), 21-41.
- Díaz-Uriarte, R., & Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*, 7(1), 1-13.
- Dudovitz, R. N., Chung, P. J., Reber, S., Kennedy, D., Tucker, J. S., Shoptaw, S., ... & Wong, M. D. (2018). Assessment of exposure to high-performing schools and risk of adolescent substance use: a natural experiment. *JAMA Pediatrics*, 172(12), 1135-1144.
- Farrington, D. P. (2010). Family influences on delinquency. *Juvenile Justice and Delinquency*, 10, 203-222.
- Gasper, J., DeLuca, S., & Estacion, A. (2012). Switching schools: Revisiting the relationship between school mobility and high school dropout. *American Educational Research Journal*, 49(3), 487-519.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. California: O'Reilly Media, Inc.
- Hirschi, T. (1969). A control theory of delinquency. In Williams III, F. P., & McShane (Ed.), *Criminology theory: Selected classic readings* (pp. 289-305). New York: Routledge.
- Hoeben, E. M., & Weerman, F. M. (2016). Why is involvement in unstructured socializing related to adolescent delinquency? *Criminology*, 54(2), 242-281.
- Hoeben, E. M., Osgood, D. W., Siennick, S. E., & Weerman, F. M. (2021). Hanging out with the wrong crowd? The role of unstructured socializing in adolescents' specialization in delinquency and substance use. *Journal of Quantitative Criminology*, 37(1), 141-177.
- Hoffmann, J. P., Thorpe, J. D., & Dufur, M. J. (2020). Family social capital and delinquent

- behavior in the United Kingdom. *Social Sciences*, 9(10), 178.
- Kierkus, C. A., & Hewitt, J. D. (2009). The contextual nature of the family structure/delinquency relationship. *Journal of Criminal Justice*, 37(2), 123-132.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28, 1-26.
- Marcal, K. E. (2021). Pathways to adolescent emotional and behavioral problems: An examination of maternal depression and harsh parenting. *Child Abuse & Neglect*, 113, 104917.
- Matsueda, R. L. (1992). Reflected appraisals, parental labeling, and delinquency: Specifying a symbolic interactionist theory. *American Journal of Sociology*, 97(6), 1577-1611.
- Miller, J. (2013). Individual offending, routine activities, and activity settings: Revisiting the routine activity theory of general deviance. *Journal of Research in Crime and Delinquency*, 50(3), 390-416.
- Muller, M. P., Tomlinson, G., Marrie, T. J., Tang, P., McGeer, A., Low, D. E., ... & Gold, W. L. (2005). Can routine laboratory tests discriminate between severe acute respiratory syndrome and other causes of community-acquired pneumonia? *Clinical Infectious Diseases*, 40(8), 1079-1086.
- Obsuth, I., Murray, A. L., Knoll, M., Ribeaud, D., & Eisner, M. (2021). Teacher-student relationships in childhood as a protective factor against adolescent delinquency up to age 17: A propensity score matching approach. *Crime & Delinquency*, 00111287211014153.
- Sabatine, E., Lippold, M., & Kainz, K. (2017). The unique and interactive effects of parent and school bonds on adolescent delinquency. *Journal of Applied Developmental Psychology*, 53, 54-63.
- Valasik, M., & Barton, M. S. (2018). The George Wilson effect: does intergenerational closure and collective efficacy reduce juvenile delinquency in a neighborhood?. *Deviant Behavior*, 39(12), 1658-1671.
- Van Ouytsel, J., Ponnet, K., & Walrave, M. (2018). Cyber dating abuse victimization among secondary school students from a lifestyle-routine activities theory perspective. *Journal of Interpersonal Violence*, 33(17), 2767-2776.
- Walters, G. D. (2018). Resistance to peer influence and crime desistance in emerging adulthood: A moderated mediation analysis. *Law and Human Behavior*, 42(6), 520.
- Warr, M. (2002). *Companions in crime: The social aspects of criminal conduct*. UK: Cambridge University Press.

논문 투고일 : 2022. 07. 27  
1 차 심사일 : 2022. 08. 09  
2 차 심사일 : 2022. 10. 12  
3 차 심사일 : 2022. 10. 26  
게재 확정일 : 2022. 10. 27

## A Study on the Classic Theory-Driven Predictors of Adolescent Online and Offline Delinquency using the Random Forest Machine Learning Algorithm

TaekHo Lee

SeonYeong Kim

YoonSun Han

Seoul National University

Adolescent delinquency is a substantial social problem that occurs in both offline and online domains. The current study utilized random forest algorithms to identify predictors of adolescents' online and offline delinquency. Further, we explored the applicability of classic delinquency theories (social learning, strain, social control, routine activities, and labeling theory). We used the first-grade and fourth-grade elementary school panels as well as the first-grade middle school panel ( $N=4,137$ ) among the sixth wave of the nationally-representative Korean Children and Youth Panel Survey 2010 for analysis. Random forest algorithms were used instead of the conventional regression analysis to improve the predictive performance of the model and possibly consider many predictors in the model. Random forest algorithm results showed that classic delinquency theories designed to explain offline delinquency were also applicable to online delinquency. Specifically, salient predictors of online delinquency were closely related to individual factors (routine activities and labeling theory). Social factors (social control and social learning theory) were particularly important for understanding offline delinquency. General strain theory was the commonly important theoretical framework that predicted both offline and online delinquency. Findings may provide evidence for more tailored prevention and intervention strategies against offline and online adolescent delinquency.

*Key words* : adolescent delinquency, online delinquency, offline delinquency, random forest algorithm, delinquency theory



## 부 록

### 비행 이론별 예측변수

이론	예측변수	척도
사회 통제 이론	학습습관 - 성취가치 평균	4점 척도 (‘1=전혀 그렇지 않다’ ~ ‘4=매우 그렇다’)
	학습습관 - 숙달목적 지향성 평균	
	학습습관 - 행동통제 평균	
	학습습관 - 학업시간 관리 평균	
	또래애착 - 의사소통 평균	
	또래애착 - 신뢰 평균	
	또래애착 - 소외 평균	
	학습활동 평균	
	학교규칙 평균	
	교우관계 평균	
	교사관계 평균	
	다문화수용성 평균	
	지역사회 인식 평균	
	공동체의식 평균	
	양육방식 - 방임 평균	
	양육방식 - 학대 평균	
일반 긴장 이론	체육시간 중 운동시간	5점 척도 (‘1=없다’ ~ ‘5=4시간 이상’)
	해외방문 연간 경험 유무	‘1=있다’, ‘0=없다’
	이성친구 유무	
	방과후 교육 연간 경험 유무	
	문화활동 연간 경험횟수	
가족/단체 여행 연간 경험횟수	횟수	
보호자의 삶의 만족도	4점 척도 (‘1=전혀 만족하지 못한다’ ~ ‘4=매우 만족한다’)	
전체 성적 만족도		

비행 이론별 예측변수

(계속 1)

이론	예측변수	척도
일반 긴장 이론	정서문제 - 주의집중 평균	4점 척도 (‘1=전혀 그렇지 않다’ ~ ‘4=매우 그렇다’)
	정서문제 - 공격성 평균	
	정서문제 - 신체증상 평균	
	정서문제 - 사회적위축 평균	
	정서문제 - 우울 평균	
	삶의 만족도 평균	4점 척도 (‘1=매우 건강하지 못하다’ ~ ‘4=매우 건강하다’)
	건강상태 평가	
	보호자의 건강상태 평가	‘1=있다’, ‘0=없다’
	비행 연간 피해경험 유무 - 심한 놀람이나 조롱	
	비행 연간 피해경험 유무 - 집단따돌림	
	비행 연간 피해경험 유무 - 폭행	
	비행 연간 피해경험 유무 - 협박	
	비행 연간 피해경험 유무 - 돈이나 물건 빼가기	
	비행 연간 피해경험 유무 - 성폭행이나 성희롱	
비행 연간 피해경험 유무 - 심한 욕설	‘1=일을 하고 있다’, ‘0=일을 하고 있지 않다’	
연간 전학 경험 유무		
부친 근로여부		
모친 근로여부	‘1=다문화가정이다’, ‘0=다문화가정이 아니다’	
다문화가정 여부		
가정의 경제수준 평가	7점 척도 (‘1=매우 못 사는 편’ ~ ‘7=매우 잘 사는 편’)	
가족구성	1=‘부모(아버지와 어머니)+자녀’, 2=‘한 부모(아버지 또는 어머니)+자녀’, 3=‘(한)조부모+자녀’, 4=‘(한)조부모+부모+자녀’, 5=‘(한)조부모+한 부모+자녀’, 6=‘기타’	

비행 이론별 예측변수

(계속 2)

이론	예측변수	척도
일반 긴장 이론	부모구성	1='친아버지+친어머니', 2='친아버지만', 3='친어머니만', 4='친아버지+새어머니', 5='친어머니+새아버지', 6='새아버지+새어머니', 7='새아버지만', 8='새어머니만', 9='부모님 모두 안 계삼'
	월평균 용돈 가구 연간 소득	만원 단위
사회 학습 이론	연간 비행행동 경험 친구수 - 담배 피우기	명 단위
	연간 비행행동 경험 친구수 - 술 마시기	
	연간 비행행동 경험 친구수 - 무단결석	
	연간 비행행동 경험 친구수 - 가출	
	연간 비행행동 경험 친구수 - 심하게 놀리거나 조롱	
	연간 비행행동 경험 친구수 - 집단따돌림	
	연간 비행행동 경험 친구수 - 패싸움	
	연간 비행행동 경험 친구수 - 폭행	
	연간 비행행동 경험 친구수 - 협박	
	연간 비행행동 경험 친구수 - 돈이나 물건 뺏기	
	연간 비행행동 경험 친구수 - 돈이나 물건 훔치기	
	연간 비행행동 경험 친구수 - 성관계	
	연간 비행행동 경험 친구수 - 성폭행이나 성희롱	
	연간 비행행동 경험 친구수 - 돈을 걸고 도박 게임하기	
형제자매 수 - 형/오빠		
형제자매 수 - 누나/언니		
형제자매 수 - 여동생		
형제자매 수 - 남동생		

비행 이론별 예측변수

(계속 3)

이론	예측변수	척도
사회 학습 이론	연간비행행동 경험 유무 - 담배 피우기	'1=있다', '0=없다'
	연간 비행행동 경험 유무 - 술 마시기	
	연간 비행행동 경험 유무 - 무단결석	
	연간 비행행동 경험 유무 - 가출	
	연간 비행행동 경험 유무 - 심하게 놀리거나 조롱	
	연간 비행행동 경험 유무 - 집단따돌림	
	연간 비행행동 경험 유무 - 패싸움	
	연간 비행행동 경험 유무 - 폭행	
	연간 비행행동 경험 유무 - 협박	
	연간 비행행동 경험 유무 - 돈이나 물건 빼기	
	연간 비행행동 경험 유무 - 돈이나 물건 훔치기	
	연간 비행행동 경험 유무 - 성관계	
	연간 비행행동 경험 유무 - 성폭행이나 성희롱	
	연간 비행행동 경험 유무 - 돈을 걸고 도박 게임하기	
	연간 사이버 비행행동 경험 유무 - 채팅/ 게시판에 거짓정보 올리기	
	연간 사이버 비행행동 경험 유무 - 불법 소프트웨어 다운로드	
	연간 사이버 비행행동 경험 유무 - 타인 아이디나 주민번호 무단 도용	
	연간 사이버 비행행동 경험 유무 - 채팅에서 성별이나 나이 속이기	
	연간 사이버 비행행동 경험 유무 - 해킹	
	연간 사이버 비행행동 경험 유무 - 채팅/게시판에서 욕이나 폭력적인 언어 사용	
형제자매 유무		

비행 이론별 예측변수

(계속 4)

이론	예측변수	척도
	휴대전화 이용빈도 - 게임 및 오락	4점 척도 (‘1=전혀 사용하지 않는다’ ~ ‘4=자주 사용한다’)
	휴대전화 이용빈도 - 사진 촬영	
	휴대전화 이용빈도 - 동영상 감상	
	휴대전화 이용빈도 - 음악 듣기	
	휴대전화 이용빈도 - 시간 보기	
	휴대전화 이용빈도 - 가족과의 통화	
	휴대전화 이용빈도 - 가족과의 문자 메시지	
	휴대전화 이용빈도 - 친구와의 통화	
	휴대전화 이용빈도 - 친구와의 문자 메시지	
일상 생활 이론	취침시간 - 등교	시 단위
	취침시간 - 비등교	
	기상시간 - 등교	
	기상시간 - 비등교	
	학원과외시간 - 등교	
	학원과외시간 - 비등교	
	학교숙제시간 - 등교	
	학교숙제시간 - 비등교	
	학원과외숙제시간 - 등교	
	학원과외숙제시간 - 비등교	
	기타공부시간 - 등교	
	기타공부시간 - 비등교	
	독서시간 - 등교	
	독서시간 - 비등교	
	컴퓨터게임오락시간 - 등교	
	컴퓨터게임오락시간 - 비등교	
	TV비디오시청시간 - 등교	
	TV비디오시청시간 - 비등교	
	친구들과 노는 시간 - 등교	
	친구들과 노는 시간 - 비등교	

비행 이론별 예측변수

(계속 5)

이론	예측변수	척도
일상 생활 이론	팬덤활동 - 선호 연예인/운동선수 유무	
	컴퓨터 사용여부	'1=있다', '0=없다'
	휴대전화 보유여부	
낙인 이론	휴대전화 종류	'1=스마트폰', '2=피쳐폰(일반휴대전화)'
	자아정체감 평균	4점 척도
	자아존중감 평균	('1=전혀 그렇지 않다' ~ '4=매우 그렇다')
기타	거주지역(자택기준)	1=서울/경기, 0=그외 지역
	부친 최종학력	'1=중학교 졸업 이하', '2=고등학교 졸업', '3=전문대 졸업', '4=대학교 졸업', '5=대학원 졸업'
	모친 최종학력	'1=중학교 졸업 이하', '2=고등학교 졸업', '3=전문대 졸업', '4=대학교 졸업', '5=대학원 졸업'
	출생년도	출생년
	교급	'1=초등학교', '2=중학교', '3=고등학교'
	성별	'1=남자', '0=여자'