

초등학생의 스마트폰 중독 관련 데이터를 활용한 머신러닝 알고리즘의 성능 분석

이 철 현 (경인교육대학교 교수)*

요 약

이 연구의 목적은 초등학생의 스마트폰 중독 관련 데이터를 활용하여 머신러닝 알고리즘의 성능을 비교 분석함으로써 최적의 머신러닝 알고리즘을 탐색하는 것이다. 이를 위하여 선행연구를 토대로 KCYPS 2018 데이터로부터 스마트폰 중독에 영향을 미치는 요인과 스마트폰 중독 성향이 포함된 데이터를 수집하였다. 수집된 데이터를 머신러닝 알고리즘인 K-최근접 이웃, 로지스틱 회귀, 서포트 벡터 머신, 의사결정트리, 랜덤 포레스트에 적용하여 산출된 머신러닝 모델의 성능을 비교 평가하였다. 그 결과 예측 변수의 상대적 중요도는 스마트폰 사용시간, 학업열의, 주의집중, 자아존중감, 우울, 공격성, 게임 사용시간의 순서로 높게 나타났다. 각 머신러닝 알고리즘의 성능은 비슷하였고, 그 중 랜덤 포레스트의 성능이 가장 좋은 것으로 나타났다. 연구 결과를 통해 얻은 시사점은 다음과 같다. 첫째, 이 연구에서 산출된 모델은 사이버 중독 예방에 대한 실과교과 수업의 방향을 적절히 안내해 주고, 실과교과의 타 영역에 대한 인공지능 적용 연구를 촉발할 것이다. 둘째, 초등학생의 스마트폰 중독 연구에 인공지능 기술 적용 가능성을 국내에서 처음으로 확인하였다는 점에서 연구의 의의를 찾을 수 있다. 이는 초등학생에 대한 스마트폰 중독 성향의 종단적 예측 모델 탐색에 기여할 수 있을 것이다. 셋째, 이 연구에서 나타난 스마트폰 중독에 영향을 미치는 요인의 상대적 중요도는 초등학생 스마트폰중독 예방을 위한 교육적 처방의 방향을 찾는 데 도움이 될 것이다.

주제어: 머신러닝, 인공지능, 스마트폰 중독, 초등학생, 예측모델

I. 서론

한국의 휴대폰 보유율은 2012년부터 크게 상승하기 시작하였고, 초등고학년의 스마트폰 보유율이 가장 크게 증가하여 2018년 기준으로 초등저학년은 37.8%, 초등고학년은 81.2%

* 교신저자 : 이철현 (leesleek@gin.ac.kr)

※ 접수일자: 2020년 11월 23일 / 심사일자 : 2020년 12월 11일 ~ 12월 25일 / 게재확정일자 : 2020년 12월 25일

에 진입하였다(정보통신정책연구원, 2019). 스마트폰은 기존 휴대폰의 기능을 넘어서 컴퓨터로 할 수 있는 거의 모든 일을 할 수 있고, 컴퓨터로 할 수 없는 일까지 할 수 있도록 함으로써 현대인의 삶을 획기적으로 변화시켰다. 스마트폰 보유율 증가와 스마트폰 기능의 확대는 스마트폰 사용을 폭발적으로 증가시켰고, 스마트폰 사용이 증가할수록 건강 문제나 사이버 범죄 등의 부정적 영향 또한 증가하고 있다. 특히, 스마트폰 중독은 초등학교생에게 부정적 영향을 미치고 있다. 코로나 19 상황에서 온라인 학습 환경에 놓여진 초등학교생들은 스마트폰을 비롯한 모바일기기를 그 이전보다 더 빈번하게 사용하게 됨으로써 스마트폰 중독의 위험에 더 크게 노출되어 있는 상황이다.

초등학생 시기는 또래와의 관계, 학교 생활 등을 통하여 사회적 기술 및 사회성을 습득하는 매우 중요한 기간으로써 초등학생의 스마트폰 중독은 학업능력의 저하, 수면 부족으로 인한 발달상의 문제, 대인관계의 어려움, 부정정서의 심화 등 부정적 양상을 유발할 수 있다(김진영, 2013; 김비, 김나미, 조현섭, 2019). 스마트폰의 과다한 사용이 스마트폰 중독의 직접적 요인임을 짐작할 수 있으나, 그에 앞서 스마트폰의 과다한 사용을 유발하는 다양한 요인은 밝히는 일이 더욱 중요함에 따라 관련된 연구(오주, 2014; 김형지, 정세훈, 2015; 류세인, 조인숙, 2015; 석말숙, 구용근, 2016; 조선희, 전경숙, 2016; 이충효, 2017; 이다솔, 김광민, 문상호, 권기현, 2019; Chou & Chou, 2019; Buctot, Kim & Kim, 2020)가 다양하게 수행되고 있다.

인공지능의 머신러닝은 분류, 군집화, 연관 관계 탐색, 이상 탐지 등을 목적으로 여러 분야에 적용되어 업무의 정확도 및 효율성을 높이고 있다. 이는 많은 데이터를 머신러닝 알고리즘에 적용 및 학습시켜 예측 모델을 구축하고, 여기에 새로운 데이터를 입력하여 결과를 예측하는 방식이다. 특정 질환 여부 예측, 주가 예측, 불량품 식별 등이 그 예이다. 이에 심리, 사회, 교육 등 여러 영역에서 특정 양상의 예측 모형이나 요인을 탐색하는 다양한 연구(조현국, 2018a; 김영식, 김훈호, 2019; 김영식, 김민석, 이길재, 2019; 고은경, 2020)가 수행되고 있다. 이러한 연구들은 특정 현상의 긍정이나 부정을 예측하는 모델을 제시함으로써 새로운 사람이 어떤 양상에 처할지 예측할 수 있게 해주어 부정적 양상을 예방할 수 있게 한다. 이와 같은 맥락에서 스마트폰 중독 성향을 예측할 수 있다면, 초등학교생의 스마트폰 중독 예방을 위해 어떤 조치를 취해야 하는지에 대한 단서를 줄 수 있을 것이다. 즉, 스마트폰 중독 성향을 예측할 수 있는 모델 탐색을 위하여 여러 머신러닝 알고리즘을 적용하여 성능을 비교 분석 및 평가해 봄으로써 초등학교생들에게 어떠한 관심을 기울이고, 처방을 해야 하는지 좀 더 과학적인 전략을 제시하는데 도움을 줄 수 있을 것이다. 이에 따라 이 연구에서는 초등학교생의 스마트폰 중독 관련 데이터를 활용하여 5개의 대표적인 머신러닝 알고리즘을 적용하여 산출된 머신러닝 모델의 성능을 비교 분석해 봄으로써 최적의 머신러닝 알고리즘을 탐색하는 것을 목적으로 하였다.

II. 이론적 배경

1. 스마트폰 중독 개념 및 중독 영향 요인

스마트폰 중독은 스마트폰을 과다하게 사용함으로써 일상생활에 지장을 줄 정도로 의존성이 커진 증상으로 볼 수 있다. 한국정보화진흥원(2011)은 스마트폰 중독의 하위요인으로 내성, 금단, 일상생활장애, 가상적 대인관계 지향성의 4가지를 제시함으로써 스마트폰의 개념을 설명하였다. 내성이란 스마트폰을 더 많이 사용해야 만족을 느끼는 증상이고, 금단은 스마트폰이 없으면 불안과 초조를 경험하는 증상이다. 일상생활장애는 스마트폰을 과다하게 사용하여 가정, 학교, 직장 등에서 문제를 일으키는 상태를 의미하고, 가상적 대인관계 지향성은 현실에서 사람들과 관계를 맺기 보다는 스마트폰을 통해 관계 맺는 것이 더 즐겁고 편한 성향을 말한다.

스마트폰 중독은 스마트폰을 과다하게 사용함으로써 발생하는 증상이나, 과다한 사용에 영향을 미치는 요인이 존재할 것이라는 예측 하에 이러한 요인을 밝히는 많은 연구가 있어왔다. 이다솔 외(2019)는 인터넷·스마트폰 과의존(중독) 선행연구를 검토하여 중독 요인을 개인요인, 가족요인, 사회요인, 인터넷 특성요인으로 구분하였다. 개인요인에는 성별, 연령, 경제수준, 심리적 요인(우울, 불안, 절망감, 자존감, 삶의 만족)이 포함되었고, 가족요인에는 가족관계, 가족응집력, 감독과 통제, 가족 갈등이 포함되었다. 사회요인에는 사회적지지, 사회적 관계, SNS, 모바일 메신저가 포함되었고, 인터넷 특성요인에는 인터넷 이용특성, 인터넷 이용시간이 포함되었다. 이충효(2017)는 인구사회관련 요인에 대한 연구, 가족관련 요인에 대한 연구, 개인관련 요인에 대한 연구 41편에 대한 메타분석을 수행하였다. 인구사회관련 요인 중 가정경제상태, 성별이 유의미한 효과크기를 가졌고, 가족관련 요인에서는 부모감독, 부모애착, 가족관계, 의사소통, 긍정적 양육태도가 유의미한 효과크기가 있는 것으로 밝혀졌다. 개인관련 변인에는 자아존중감, 자기통제력, 충동성, 공격성, 자기애, 불안, 우울, 회피적 정서조절, 자극추구, 스트레스 대처방식, 스트레스, 사회적 위축의 효과크기가 유의미한 것으로 나타났다. 오주(2014)의 연구에서는 스마트폰 중독에 영향을 미치는 요인이 공격성, 스마트폰 사용횟수, 자존감, 충동성, 부모관심, 외로움, 부모통제 순으로 나타났다. 석말숙 외(2016)의 연구에 따르면 개인요인 중 하위요소인 자기통제력, 자아존중감, 가족요인 중 하위요소인 부모지지와 해결지향형 의사소통이 스마트폰 중독에 부정적인 영향력을 미치는 것으로 나타났다.

이상과 같은 선행연구에서 제시한 스마트폰 중독 요인을 고려하여, 이 연구에서 사용한 한국아동·청소년패널조사(KCYPS 2018)(하형석, 김성은, 이용해, 2019)의 데이터의 조사 항목 중 스마트폰 중독 요인에 해당하는 것을 선별하였다. 그 결과, 자아존중감, 주의집중, 공격성, 우울, 협동심, 사회적 위축, 학업열의, 스마트폰 사용시간, 컴퓨터 사용시간, 게임 이용시간의 10개를 선별할 수 있었다.

2. 머신러닝

머신러닝은 컴퓨터가 데이터로부터 특정 문제해결을 위한 지식을 자동으로 추출해서 사용할 수 있게 하는 기술(이건명, 2018)로써 인공지능 기술의 주요 영역이다. 머신러닝은 어떤 상황과 그 상황에 따른 결과를 조사한 데이터로부터 패턴을 찾아내어, 새로운 상황에 대한 결과를 예측하도록 한다. 예를 들어, 특정 질환을 가진 환자들의 신체적 특성 데이터를 학습시켜 패턴을 찾아내고, 새로운 사람들의 신체적 특성 데이터를 그 패턴에 적용하여 그 사람의 특정 질환 여부를 예측할 수 있도록 한다.

머신러닝은 크게 지도학습(supervised learning), 비지도학습(non-supervised learning), 강화학습(reinforcement learning)으로 분류된다. 지도학습은 학습 데이터에 입력값(예측 변수)과 출력값(종속 변수)이 함께 있을 때의 학습방법이고, 비지도학습은 학습 데이터에 입력값만 있을 때의 학습방법이다. 강화학습은 특정 환경에서 보상과 벌칙을 통해 큰 보상을 받기 위해 최상의 전략을 스스로 학습하는 방법이다(김의중, 2016; 오렐리앙 제롱, 2018). 예를 들어, 사람들의 신체적 특성값과 질환여부값이 함께 있는 데이터는 지도학습을 적용하게 되고, 소비자의 소비행동 특성값만 있는 데이터는 비지도학습을 적용할 수 있다. 강화학습의 대표적 사례로는 인공지능 바둑 알고리즘인 알파고를 들 수 있다.

머신러닝의 지도학습과 비지도학습의 주요 알고리즘을 제시하면 <표 1>과 같다. 여기에서 지도학습은 주로 분류를 목적으로 하는 방법이고, 비지도학습은 군집화, 특성 추출, 이상치 탐지, 연관 규칙 탐색 등을 목적으로 한다. 본 연구에서 사용하는 데이터는 학생들의 여러 특성값을 예측 변수로 하고, 스마트폰 중독 여부를 종속 변수로 두기 때문에 머신러닝 중 지도학습을 적용하고자 한다.

<표 1> 머신러닝의 지도학습과 비지도학습의 주요 알고리즘

분류		주요 알고리즘
지도학습		k-최근접 이웃(k-nearest neighbors) 로지스틱 회귀(logistic regression) 서포트 벡터 머신(support vector machine) 의사결정트리(decision tree) 랜덤 포레스트(random forests) 선형 회귀(linear regression)
비지도학습	군집	k-평균(k-means) 계층 군집 분석(hierarchical cluster analysis) 기댓값 최대화(expectation maximization)
	시각화와 차원 축소	주성분 분석(principal component analysis) 커널 PCA 지역적 선형 임베딩(locally-linear embedding)
	연관 규칙 학습	어프라이어리(apriori) 이클랫(eclat)

Ⅲ. 연구 방법

1. 연구 대상

본 연구는 2018년에 실시된 한국아동·청소년패널조사(KCYPS 2018)(하형석 외, 2019) 자료를 사용하였다. 이 자료는 초등학교 4학년 코호트 2,607명, 중학교 1학년 코호트 2,590명을 원패널로 하였고, 원패널의 보호자/형제자매를 각각 2,500여명, 1,000여명씩 포함하여 조사된 것이다. 아동을 대상으로 개인발달(생활시간, 지적발달, 진로, 사회/정서/역량발달, 비행, 건강, 신체발달 등), 발달환경(가정, 학교, 친구, 매체, 활동/문화 환경 등) 등을 조사 내용으로 하였고, 보호자를 대상으로는 배경변인(학력, 직업, 경제수준 등), 보호자 개인 특성(사회/정서/역량, 매체이용, 건강 등), 아동·청소년 발달 배경(가정, 교육 등)을 조사내용으로 하였고, 응답은 4점 척도를 이용하였다.

본 연구에서는 2018 한국아동·청소년패널조사 자료 중 초등학교 4학년 원패널 2,607명을 연구의 대상으로 하였고, 결측치가 있는 209명의 데이터를 제외하고, 2398명의 자료를 최종 분석에 사용하였다. 분석에 사용한 자료의 거주지와 성별 변인의 빈도 분석 결과는 <표 2>와 같다.

<표 2> 분석 자료의 일반 변인

구분		빈도	퍼센트
거주지역	서울	365	15.2
	부산	136	5.7
	대구	106	4.4
	인천	145	6.0
	광주	77	3.2
	대전	58	2.4
	울산	55	2.3
	세종	28	1.2
	경기	616	25.7
	강원	98	4.1
	충북	90	3.8
	충남	133	5.5
	전북	94	3.9
	전남	80	3.3
	경북	132	5.5
	경남	158	6.6
	제주	27	1.1
	전체	2398	100.0
성별	남자	1192	49.7
	여자	1206	50.3
	전체	2398	100.0

2. 측정 도구

가. 예측 변수

이 연구에서 아동의 스마트폰 중독을 예측하는 변수는 스마트폰 중독 영향 요인의 선행연구와 이 연구에서 사용한 데이터를 토대로 자아존중감, 주의집중, 공격성, 우울, 협동심, 사회적 위축, 학업열의, 스마트폰 사용시간, 컴퓨터 사용시간, 게임 이용시간의 10개로 설정하였다.

자아존중감은 자신에 대한 만족, 용도, 장점, 가치, 자랑스러움, 쓸모, 역량, 존중, 실패자 경향, 긍정적 태도 등 8개의 문항으로 구성되었고, ‘전혀 그렇지 않다(1)’, ‘그렇지 않은 편이다(2)’, ‘그런 편이다(3)’, ‘매우 그렇다(4)’의 4점 리커트 척도로 응답하여 평균 점수가 높을수록 자아존중감이 높은 것으로 해석된다. 주의집중은 주의 산만, 문제 읽기, 학용품 분실, 글자 쓰기 등 7개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 주의집중력이 높은 것으로 해석된다. 공격성은 남의 일 방해, 작은 일 트집 등 5개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 공격성이 높은 것으로 해석된다.

우울은 걱정, 불행 자각, 울기 정도, 외로움 등 10개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 우울이 심한 것으로 해석된다. 협동심은 친구의 장점 인지시키기, 친구의 어려움 도움, 자유로운 이야기 분위기 조성, 새로운 그룹활동 방법 제시, 자신의 역할 최선을 다하기 등 14개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 협동심이 높은 것으로 해석된다.

사회적 위축은 어색함, 부끄러움, 수줍음 등 5개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 사회적 위축이 심한 것으로 해석된다. 학업열의는 공부의 의미와 목적 인식, 공부의 뿌듯함, 자아실현 인식, 학교 등교 의욕, 공부 집중 등 16개 문항에 대하여 자신에게 해당하는 정도를 4점 리커트 척도로 응답하여 평균 점수가 높을수록 학업열의가 높은 것으로 해석된다.

스마트폰 사용시간, 컴퓨터 사용시간은 평일과 주말에 스마트폰, 컴퓨터를 가지고 노는 시간을 ‘전혀 안함(1)’, ‘30분 미만(2)’, ‘30분 ~ 1시간 미만(3)’, ‘1시간 ~ 2시간 미만(4)’, ‘2시간 ~ 3시간 미만(5)’, ‘3시간 ~ 4시간 미만(6)’, ‘4시간 이상(7)’의 7점 리커트 척도로 응답한 결과의 평균값으로 점수가 높을수록 스마트폰 사용시간, 컴퓨터 사용시간이 많은 것으로 해석된다. 게임 이용시간은 스마트폰의 사용 목적 중 게임 이용에 대하여 ‘전혀 사용하지 않는다(1)’, ‘거의 사용하지 않는다(2)’, ‘가끔 사용한다(3)’, ‘자주 사용한다(4)’의 4점 리커트 척도로 응답하여 점수가 높을수록 게임 이용시간 이 많은 것으로 해석된다.

머신러닝 분석에서 사용된 예측 변수와 종속 변수의 기술 통계 분석 결과는 <표 3>과 같다.

<표 3> 예측 변수와 종속 변수의 기술 통계 결과

	변인	N	평균	최소값	최대값	표준편차
예측 변수	자아존중감	2398	3.21	1	4	0.48
	주의집중	2398	2.98	1	4	0.57
	공격성	2398	1.76	1	4	0.57
	우울	2398	1.54	1	4	0.53
	협동심	2398	2.99	1	4	0.54
	사회적 위축	2398	2.03	1	4	0.74
	학업열의	2398	2.72	1	4	0.54
	스마트폰 사용시간	2398	3.69	1	4	1.43
	컴퓨터 사용시간	2398	1.69	1	7	1.10
	게임 이용 시간	2398	3.12	1	4	0.95
종속 변수	스마트폰 중독	2398	1.81	1.00	4.00	0.51

나. 종속 변수

이 연구에서 종속 변수는 스마트폰 중독 척도의 채점에 의해 분류된 3개의 사용자군을 이항변수로 변형하여 사용하였다. 스마트폰 중독 척도는 한국정보화진흥원(2011)의 청소년 스마트폰 중독 자가진단 척도 15문항을 이용한 것으로 일상생활장애(1요인) 5문항, 가상세계지향(2요인) 2문항, 금단(3요인) 4문항, 내성(4요인) 4문항으로 구성되어 있고, 각 문항에 대한 응답은 4점 리커트 척도로 구성되어 있다. 스마트폰 중독 척도를 적용하여 얻은 응답자의 원점수는 <표 4>의 기준에 따라 채점하여 고위험사용자군, 잠재적위험사용자군, 일반사용자군의 3개 집단으로 구분된다.

<표 4> 청소년 스마트폰 중독 척도 채점 기준

채점 방법	[1단계] 문항별	전혀 그렇지 않다 : 1점, 그렇지 않다 : 2점, 그렇다 : 3점, 매우 그렇다 : 4점 ※ 단, 문항 8번, 10번, 13번은 다음과 같이 역채점 실시 <전혀 그렇지 않다 : 4점, 그렇지 않다 : 3점, 그렇다 : 2점, 매우 그렇다 : 1점>
	[2단계] 총점 및 요인별	총 점 ▶ ① 1~15번 합계 요 인 별 ▶ ② 1요인 일상생활장애(1,5,9,12,13번) 합계 ③ 3요인 금단(3,7,10,14번) 합계 ④ 4요인 내성(4,8,11,15번) 합계
고위험 사용자군	총 점 ▶ ① 45점 이상 요인별 ▶ ② 1요인 16점 이상 ③ 3요인 13점 이상 ④ 4요인 14점 이상 판정 : ①에 해당하거나, ②~④ 모두 해당되는 경우	
잠재적 위험 사용자군	총 점 ▶ ① 42점 이상~44점 이하 요인별 ▶ ② 1요인 14점 이상 ③ 3요인 12점 이상 ④ 4요인 13점 이상 판정 : ①~④ 중 한 가지라도 해당되는 경우	
일반 사용자군	총 점 ▶ ① 41점 이하 요인별 ▶ ② 1요인 13점 이하 ③ 3요인 11점 이하 ④ 4요인 12점 이하 판정 : ①~④ 모두 해당되는 경우	

위 채점기준에 따라 이 연구에서 2018 한국아동·청소년패널조사 자료 중 초등학교 4학년 2,398명의 스마트폰 중독 점수를 채점한 결과, 고위험사용자군 48명, 잠재적위험사용자군 307명, 일반사용자군 2043명이라는 결과를 얻게 되었다. 이 중 고위험사용자군의 빈도가 전체 연구대상자수에 비하여 상대적으로 너무 적어서 머신러닝을 적용하는 분석 빈도로써 적합하지 않다고 판단하여, 고위험사용자군과 잠재적위험사용자군을 합쳐서 위험사용자군, 일반사용자군의 이항변수로 변형하여 머신러닝 분석에 투입하였다.

3. 분석 과정

이 연구에서는 잠정적 예측 변수 중 종속 변수에 영향을 미치는 변수를 밝히기 위하여 다중회귀분석을 사용하였고, 머신러닝 모델 구축을 위하여 대표적인 지도학습 방법인 K-최근접 이웃, 로지스틱 회귀, 서포트 벡터 머신, 의사결정트리, 랜덤 포레스트를 차례대로 적용하였다.

K-최근접 이웃은 각 데이터들 간에 거리를 측정하여 가까운 k개의 다른 데이터의 레이블을 참조하여 분류하는 방법으로 최적의 k를 찾는 것이 주요 과제이다. 로지스틱 회귀는 종속변수가 범주형 자료일 경우 적용하는 회귀분석으로 판별분석과 유사하나 독립변수의 정규분포를 엄격히 가정하지는 않는다. 서포트 벡터 머신은 주어진 데이터 카테고리를 기준으로, 새로운 데이터가 어느 카테고리에 속할지 판단하는 분류모델로 뛰어난 성능으로 딥러닝 이전에 가장 활용도가 높았던 분류 모델에 속한다. 의사결정 트리는 각 변수의 영역을 반복적으로 분할함으로써 전체 영역에서의 규칙을 생성하고, 의사결정 규칙을 트리 구조로 도표화하여 관심대상이 되는 집단을 몇 개의 소집단으로 분류하는 방법이다. 랜덤 포레스트는 학습 데이터로 여러 의사결정트리를 구성하여 분석하고, 이를 종합하는 앙상블(ensemble) 기법으로 학습 데이터를 무작위로 샘플링해서 다수의 의사결정트리 분석을 하기 때문에 높은 정밀도를 보인다(김원표, 서정용, 김정규, 2018; 이견명, 2018; 오렐리앙 제룽, 2018).

머신러닝을 적용한 모델 탐색을 위하여 파이썬(python) 패키지 관리 시스템인 아나콘다(anaconda)를 사용하였다. 아나콘다 패키지에 포함되어 있는 주피터 노트북(jupyter notebook) 환경에서 파이썬과 파이썬용 머신러닝 라이브러리인 사이킷런(scikit learn)을 이용하여 모델을 코딩하고 테스트하였다. 또한, 데이터 조작 및 그래프 작성을 위하여 판다스(pandas) 및 매트플롯라이브(matplotlib) 등의 파이썬 라이브러리를 사용하였다.

머신러닝 모델링은 데이터를 훈련 세트(train set)와 테스트 세트(test set)로 나눈 다음, 훈련 세트를 사용해 모델을 훈련시키면서 최적화된 하이퍼파라미터를 찾아내고, 테스트 세트를 사용해 모델을 평가함으로써 새로운 샘플에 모델이 얼마나 잘 작동할지 예측하게 해준다. 보통 데이터의 80%를 훈련에 사용하고 20%는 테스트에 사용하나, 훈련 세트와 테스트 세트의 비율을 70:30 또는 75:25의 비율로 정할 수도 있다(오렐리앙 제룽, 2018).

본 연구에서는 사이킷런 라이브러리의 `train_test_split` 모듈의 기본설정인 훈련 세트 75%, 테스트 세트 25%의 비율로 데이터를 나누어 사용하였다.

하이퍼파라미터는 머신러닝 모델의 성능을 제어하기 위하여 설정하는 변인으로 머신러닝 알고리즘에 따라 다르고, 사용자가 직접 설정한다(Kuhn & Johnson, 2013). 머신러닝 알고리즘 내에서 하이퍼파라미터가 달라짐에 따라 주어진 과제에 가장 적합한 정확도와 일반화 모델이 존재하므로, 그리드 서치(grid search) 및 랜덤 서치(random search)등 모델에 대한 세부 튜닝 과정을 거쳐야 한다(김원표 외, 2018). 본 연구에서는 하이퍼파라미터를 직접 지정하는 그리드 서치 방법과 지정한 범위 내에서 무작위로 하이퍼파라미터를 선정하여 최적의 하이퍼파라미터를 자동으로 찾아주는 랜덤 서치 방법을 사용하였다. 각 머신러닝 알고리즘의 주요 하이퍼파라미터와 그에 대한 개념을 제시하면 <표 5>와 같다.

<표 5> 머신러닝 알고리즘의 주요 하이퍼파라미터

알고리즘	주요 하이퍼파라미터
k-최근접 이웃	• <code>n_neighbors</code> : 이웃 케이스의 수(K)
로지스틱 회귀	• <code>C</code> : 기준값은 1이며 값이 작을수록 모델이 단순해지고, 값이 커질수록 모델이 복잡해짐
서포트 벡터 머신	• <code>C</code> : 기준값은 1이며 값이 작을수록 모델이 단순해지고, 값이 커질수록 모델이 복잡해짐 • <code>gamma</code> : 하나의 훈련 샘플에 미치는 영향을 의미하며 이 값이 작으면 넓은 영역, 크면 영역이 제한적임
의사결정트리	• <code>max_depth</code> : 트리의 최대 깊이 • <code>min_samples_leaf</code> : 리프 노드에 있어야 하는 최소 샘플의 수
랜덤 포레스트	• <code>max_features</code> : 최적 분할에 사용되는 특성 • <code>n_estimators</code> : 트리의 수 (디폴트:10)

머신러닝 모델의 성능 평가는 출력값이 2개 이상의 클래스가 될 수 있는 머신러닝 성능 측정에 사용되는 혼동행렬(confusion matrix)을 사용하였다. 혼동행렬은 [그림 1]과 같이 종속 변수의 실제와 예측의 긍정(Positive), 부정(Negative)에 따라 매칭되는 4가지 경우의 수를 테이블 형태로 나타낸 것이다.

		실제	
		Positive	Negative
예측	Positive	True Positive(TP)	False Positive(FP)
	Negative	False Negative(FN)	True Negative(TN)

[그림 1] 혼동행렬

혼동행렬을 통해 머신러닝 분류 모델의 성능을 평가할 수 있는 대표적인 지표는 정확도(accuracy), 재현율(recall), 정밀도(precision), f1-score이다. 재현율은 실제 긍정인 것들 중

에서 긍정으로 예측한 비율($TP/(TP+FN)$)이고, 정밀도는 긍정으로 예측한 것들 중에서 실제 긍정인 경우의 비율($TP/(TP+FP)$)이다. f1-score는 재현율과 정밀도가 가진 한계를 보완한 것으로 재현율과 정밀도의 조화평균값($2*(\text{재현율}*\text{정밀도})/(\text{재현율}+\text{정밀도})$)이다.

이상과 같은 분석 과정은 데이터 분할, 모델 적용, 하이퍼파라미터 탐색 및 결정, 모델 성능 평가의 순서로 요약할 수 있다.

IV. 연구 결과

1. 스마트폰 중독 성향에 대한 예측 변수

학습자의 스마트폰 중독 성향에 대한 예측 변수 중 스마트폰 중독 성향에 통계적으로 유의한 영향을 미치는 변수를 선정하기 위하여 머신러닝의 다중회귀분석을 수행하였다. 다중회귀분석은 종속 변수가 연속형 자료일 경우 적용할 수 있는 지도학습 방법이다. 즉, 예측 변수에는 자아존중감, 주의집중, 공격성, 우울, 협동심, 사회적 위축, 학업열의, 스마트폰 사용시간, 컴퓨터 사용시간, 게임 이용시간의 10개 자료를 설정하였고, 종속 변수에는 스마트폰 중독 평균값을 설정하였다. 스마트폰 중독 성향에 대한 다중회귀분석 결과는 <표 6>과 같다.

<표 6> 스마트폰 중독 성향에 대한 다중회귀분석 결과

R 제곱	수정된 R 제곱	Durbin-Watson	p	
.452	.442	2.061	.000	

예측 변수	결정계수	표준편차	t	p
자아존중감	-.1888	.047	-4.020**	.000
주의집중	-.1766	.039	-4.587**	.000
공격성	.1339	.039	3.452**	.001
우울	-.0909	.044	-2.044*	.041
협동심	.0105	.035	.302	.763
사회적 위축	-.0052	.025	-.205	.837
학업열의	-.1156	.038	-3.023**	.003
스마트폰 사용시간	.1266	.013	9.946**	.000
컴퓨터 사용시간	-.0021	.015	-.144	.886
게임 사용시간	.0583	.018	3.264**	.001

*p < .05, **p < .01

분석 결과, Durbin-Watson값은 2.061로써 2에 근사하므로 독립 변인들 사이의 상관관계

는 무시할 수 있는 수준이었다. 예측 변수 중 5% 유의수준에서 통계적으로 유의한 결과 보인 것은 자아존중감, 주의집중, 공격성, 사회적 위축, 학업열의, 스마트폰 사용시간, 게임 이용시간의 7개로 나타나 이 7개의 예측 변수를 머신러닝에 적용하는 것으로 결정하였다.

2. 머신러닝 모델 성능 비교

연구데이터의 자아존중감, 주의집중, 공격성, 사회적 위축, 학업열의, 스마트폰 사용시간, 게임 이용시간을 예측 변수로 하고, 스마트폰 중독에 대한 위험사용자군, 일반사용자군의 이항변수를 종속 변수로 하여 K-최근접 이웃, 로지스틱 회귀, 서포트 벡터 머신, 의사결정트리, 랜덤 포레스트 5가지 알고리즘을 적용한 머신러닝을 수행하였다. 알고리즘별로 우수한 성능의 모델을 찾기 위하여 그리드 서치 방법과 랜덤 서치 방법을 반복적으로 여러 차례 적용하며, 최적의 하이퍼파라미터를 탐색하였다.

머신러닝 모델의 성능 평가를 위하여 사이킷런 라이브러리의 metrics 패키지에서 제공하는 classification_report 함수를 이용하여 정확도, 정밀도, 재현율, f1-score를 산출하였다. 이 때 종속 변수인 이분 변수의 각 클래스에 해당하는 데이터의 개수에 가중치를 주어 평균을 계산한 가중치평균(weighted average)값을 사용하였다. 이와 같은 방법에 따라 산출된 머신러닝 모델의 성능 평가 결과를 제시하면 <표 7>과 같다.

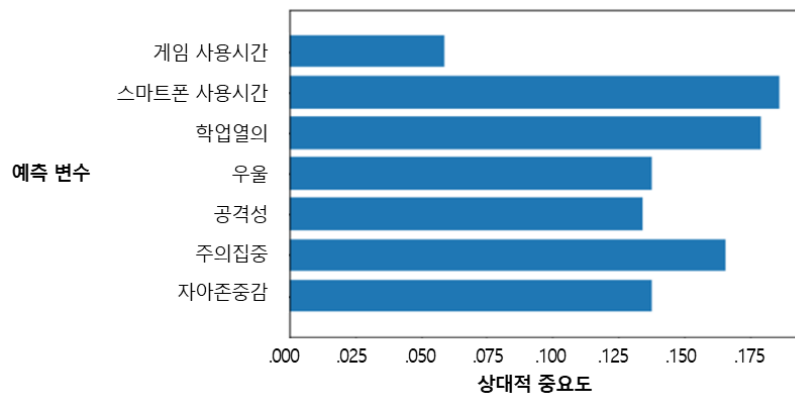
<표 7> 머신러닝 모델 성능 비교

알고리즘	정확도	정밀도	재현율	f1-score	하이퍼파라미터 값
K-최근접 이웃	.78	.84	.78	.81	n_neighbors = 2
로지스틱 회귀	.86	.83	.86	.83	C = 72
서포트 벡터 머신	.86	.81	.86	.80	C = 43 gamma = 16 kernel = 'rbf'
의사결정트리	.87	.84	.87	.83	max_depth = 11 min_samples_leaf = 48
랜덤 포레스트	.87	.85	.87	.84	max_features = 'log2' n_estimators = 187

정밀도의 경우, K-최근접 이웃, 로지스틱 회귀, 서포트 벡터 머신, 의사결정트리, 랜덤 포레스트가 모두 비슷한 수준으로 나타났고, 그 중 랜덤 포레스트의 정밀도가 0.85로 가장 높았다. 실제 일반사용자군 중에서 일반사용자군으로 예측한 비율인 재현율의 경우, K-최근접 이웃이 0.78로 다른 4개의 알고리즘에 비하여 상대적으로 낮게 나타났고, 나머지 알고리즘의 성능은 비슷하였다. f1-score를 기준으로 했을 때, 5개 알고리즘이 각각 0.81, 0.83, 0.80, 0.83, 0.84로 나타나 성능이 비슷한 것으로 나타났고, 그 중 랜덤 포레스

트의 f1-score가 0.84로 가장 성능이 좋은 것으로 나타났다.

가장 우수한 성능을 보인 랜덤 포레스트 모델에서 matplotlib를 이용하여 산출한 예측 변수의 상대적 중요도를 제시하면 [그림 2]와 같다. 예측 변수의 상대적 중요도는 스마트폰 사용시간(.185), 학업열의(.179), 주의집중(.166), 자아존중감(.138), 우울(.137), 공격성(.133), 게임 사용시간(.059)의 순서로 높게 나타났다.



[그림 2] 랜덤 포레스트 모델에서 예측 변수의 상대적 중요도

여기서의 상대적 중요도는 랜덤 포레스트 알고리즘의 반복적인 데이터학습을 통해 최적의 하이퍼파라미터를 탐색한 결과로 산출된 것이므로 데이터의 특성을 충분히 반영한 결과로 볼 수 있다. 또, 의사결정트리는 학습데이터에 따라 트리구조가 달라지기 때문에 일반화하여 사용하는데 안정적이지 못한 반면, 랜덤 포레스트는 일반화 성능을 향상시켜 정확도와 안정성이 높다는 특징(김원표 외, 2018) 때문에 예측 변수의 상대적 중요도는 믿을만한 것으로 볼 수 있다. 회귀분석과 같이 전통적 통계분석은 데이터를 통해 현재 상태를 진단하는 것에 그치는 반면, 머신러닝을 통해 구축된 모델은 새로운 예측 변수의 데이터만 입력하면, 종속 변수의 값을 예측할 수 있다는 데에서 의의를 찾을 수 있다.

3. 논의

이상과 같은 연구결과를 통해 얻을 수 있었던 시사점 및 연구결과와 선행연구 간 비교 분석을 중심으로 이 연구의 의의를 다음과 같이 제시하고자 한다.

첫째, 초등학교의 스마트폰 중독은 2015개정 초등학교 실과교육과정 기술 활용 영역의 ‘[6월05-05] 사이버 중독 예방, 개인 정보 보호 및 지식 재산 보호의 의미를 알고 생활 속에서 실천한다.’라는 성취기준을 통해 다루어질 수 있다. 따라서 스마트폰 중독 성향을 예측하기 위하여 머신러닝을 적용하여 각 알고리즘의 성능을 비교분석하는 작업은 실과

교과를 통해 사이버 중독 예방에 대한 교육과정을 실천하는 데에 도움을 줄 수 있을 것이다. 또한, 실과교과의 주요 영역 중 하나인 사이버 중독 예방과 관련한 새로운 연구 방법론을 개척하고, 실과교과의 타 영역에 인공지능 기법을 적용하는 연구를 수행하는데 도움을 줄 수 있을 것으로 기대한다.

둘째, 이 연구는 초등학생의 스마트폰 중독 성향을 예측하는 모델을 탐색하는 데에 인공지능 기술 적용 가능성을 국내에서 처음으로 확인하였다는 점에서 연구의 의의를 찾을 수 있을 것이다. 최근까지 머신러닝을 적용하여 과학의 미적 가치 인식, 학습자 성취, 독서활동, 사교육 참여, 대학 중도탈락률, 아이돌 생존 가능성, 행복감 등을 예측하는 연구(조현국, 2018a; 조현국, 2018b; 김영식 외, 2019; 김영식, 김민석 외, 2019; 고은경, 2020; 이정은, 김다솜, 조일현, 2020; 이은정, 송영수, 김지하, 오수현, 2020; 김슬아, 안주혁, 최복권, 2020; 이창묵, 이현숙, 나우열, 2020)가 수행되어 왔으나, 머신러닝을 적용하여 스마트폰 중독 성향을 예측한 연구는 아직까지 찾아보기 어렵다. 이 연구에서 사용한 한국아동·청소년패널조사 자료는 2018년 이후 매년 종단적으로 수집되어 공개되기 때문에 새롭게 수집된 자료를 이용하여 초등학생에 대한 스마트폰 중독 성향의 종단적 예측 모델을 탐색하는 데에 이 연구의 결과가 활용될 수 있을 것으로 기대할 수 있다.

셋째, 스마트폰 중독에 영향을 미치는 요인의 상대적 중요도가 순서대로 스마트폰 사용 시간, 학업열의, 주의집중, 자아존중감, 우울, 공격성, 게임 사용시간임이 밝혀진 것은 초등학생의 스마트폰중복 예방을 위하여 어떠한 교육적 처방을 해야 하는지를 시사한다. 스마트폰 사용시간의 상대적 중요도가 가장 높게 나타난 것은 스마트폰 중독의 여러 변인들 중에서 일평균 스마트폰 사용시간의 상대적 중요도가 가장 높은 것으로 나타났다고 보고한 주석진, 조성심(2015)의 연구와 맥을 같이 하는 결과이다. 나머지 중독 요인 또한 선행연구(오주, 2014; 석말숙, 구용근, 2016; 이충효, 2017; 이다솔 외, 2019)를 통해 그 근거를 찾을 수 있다.

V. 결론 및 제언

이 연구는 초등학생의 스마트폰 중독 예방을 위해 좀 더 과학적인 접근을 찾기 위한 방안으로 초등학생의 스마트폰 중독 관련 데이터에 머신러닝 알고리즘을 적용하여 산출된 머신러닝 모델의 성능을 비교 분석해 봄으로써 최적의 머신러닝 알고리즘을 탐색하는 것을 목적으로 하였다. 연구목적 달성을 위하여 한국아동·청소년패널조사(KCYPS 2018)(하형석 외, 2019) 자료를 토대로 다섯 개의 머신러닝 알고리즘을 적용하여 산출된 머신러닝 모델의 성능을 비교 평가하였다. 연구 결과와 앞에서 제시한 논의를 토대로 얻은 결론은 다음과 같다.

첫째, 초등학생의 스마트폰 중독 관련 데이터에 머신러닝에서 주로 사용되는 5개 알고

리즘인 K-최근접 이웃, 로지스틱 회귀, 서포트 벡터 머신, 의사결정트리, 랜덤 포레스트를 적용한 결과 각 알고리즘의 성능은 비슷하였으나, 그 중 랜덤 포레스트의 성능이 가장 우수한 것으로 나타났다. 이러한 연구 결과는 각 알고리즘별로 최적의 하이퍼파라미터 값 산출을 위하여 랜덤 서치를 여러 차례 수행하여 얻어진 것이다. 즉 하이퍼파라미터 값에 따라 각 알고리즘의 성능이 조금씩 달라지므로 랜덤 서치를 할 때마다 다르게 나오는 하이퍼파라미터 값을 계속 알고리즘에 적용해 가면서 각 알고리즘마다 가장 좋은 모델 평가 결과가 나올 때까지 머신러닝을 수행해야 얻을 수 있는 결과인 것이다. 이러한 과정을 통해 랜덤 포레스트의 성능이 가장 우수한 것으로 확인하였기 때문에 추후 초등학교생의 스마트폰 중독과 관련한 새로운 예측 변수값이 조사되었을 때, 그 값을 이 연구에서 산출된 랜덤 포레스트 모델에 적용하여 해당 데이터에 대한 스마트폰 중독 여부를 예측해 보는 일이 가능할 것이다.

둘째, 예측 변수의 상대적 중요도는 스마트폰 사용시간, 학업열의, 주의집중, 자아존중감, 우울, 공격성, 게임 사용시간의 순서로 높게 나타났다. 이러한 연구 결과는 초등학교생의 스마트폰 중독 예방을 위하여 스마트폰 사용시간을 적절히 조절하고, 학업의 흥미를 유발 및 유지하며, 집중력을 높일 수 있는 교육적 접근이 필요함을 시사한다. 또한, 초등학교생의 자아존중감을 높이고, 우울 성향과 공격성을 줄이기 위한 처방에 좀 더 노력해야 함을 알려준다. 게임 사용시간의 상대적 중요도가 가장 낮게 나타난 점은 다소 의외의 결과이나, 초등학교생의 스마트폰 사용시간에서 게임 사용시간이 차지하는 비중이 상대적으로 적다는 것을 반증하는 결과이기도 하다. 그럼에도 게임 사용시간이 스마트폰 중독을 예측하는 변수 중 하나이므로 게임 사용시간 또한 적절히 조절될 필요가 있다.

이 연구를 통해 연구의 제한점을 극복하기 위한 방안과 후속 연구를 위하여 다음과 같은 사항을 제언한다.

첫째, 이 연구에서 사용한 데이터의 훈련 세트와 테스트 세트의 비율을 75:25가 아닌 60:40, 70:30, 90:10 등으로 구분하여 적용시키고, 최적의 하이퍼파라미터값을 탐색하기 위한 그리드 서치 및 랜덤 서치를 반복적으로 수행해 봄으로써 좀 더 성능이 좋은 머신러닝 알고리즘 모델을 찾으려는 시도가 필요하다.

둘째, 실과교과 내용과 관련된 공공데이터를 탐색 및 발굴하고, 홍보 및 공유를 위한 노력이 필요하다. 이는 실과교과를 구성하는 가정영역 및 기술영역의 여러 학습 분야에 인공지능 기술을 적용하는 연구를 가능하게 함으로써 아동의 실생활 문제 해결과 실과 교수학습 방법을 개선하는데 기여할 수 있을 것이다.

셋째, 이 연구에서 이용한 다섯 가지 머신러닝 알고리즘 외에 더 많은 알고리즘을 이용하여 모델을 탐색하는 시도가 필요하다. 아울러, 긍정과 부정의 분류를 위한 알고리즘 외에 스마트폰 중독 성향을 군집화(clustering)할 수 있는 k-평균, 계층 군집 분석 등의 머신러닝 알고리즘이나 딥러닝을 적용하여 초등학교생의 스마트폰 중독 영역에 좀 더 심화된 인공지능 적용 연구가 시도될 필요가 있다.

참고 문헌

- 고은경(2020). 한국 초등학교 3학년 아동의 행복감 예측을 위한 머신러닝 알고리즘의 적용. **학습자중심교과교육연구**, 20(13), 1113-1128.
- 김비, 김나미, 조현섭(2019). 초등학생의 스마트폰 중독이 학교생활적응에 미치는 영향 : 긍정심리자본과 자기통제력의 매개효과. **알코올과 건강행동학회**, 20(1), 55-68.
- 김슬아, 안주혁, 최복권(2020). 머신러닝 기법을 활용한 아이돌 생존 가능성 예측 연구: 산업 경쟁력 증진을 중심으로. **한국콘텐츠학회논문지**, 20(5), 291-302.
- 김영식, 김민석, 이길재(2019). 랜덤 포레스트를 활용한 고교생의 독서활동 예측 요인 분석. **교육재정경제연구**, 28(4), 137-156.
- 김영식, 김훈호(2019). 머신러닝 기법을 활용한 사교육 참여 예측 모형 탐색. **교육재정경제연구**, 28(3), 29-52.
- 김원표, 서정용, 김정규(2018). **Python을 활용한 머신러닝**. 와이즈인컴퍼니.
- 김의중(2016). **인공지능, 머신러닝, 딥러닝 입문**. 위키북스.
- 김진영(2013). 초등학생의 스마트폰 중독이 학교적응유연성에 미치는 영향. **한국교류분석상담학회**, 3(2), 63-79.
- 김형지, 정세훈(2015). 초등학생의 스마트폰 중독에 영향을 미치는 요인에 관한 연구: 부모중재, 학교교육 및 심리적 반발심을 중심으로. **사이버커뮤니케이션학보**, 32(1), 87-120.
- 류세인, 조인숙(2015). 초등학생의 스마트폰 중독에 미치는 영향요인. **한국산학기술학회**, 16(9), 6180-6189.
- 석말숙, 구용근(2016). 청소년의 개인요인, 가족요인, 학교요인이 스마트폰 중독에 미치는 영향. **청소년복지연구**, 18(1), 53-77.
- 오렐리앙 제롱 지음, 박해선 옮김(2018). **Hands-On Machine Learning with Scikit-Learn & TensorFlow. 핸즈온 머신러닝**. 한빛미디어.
- 오주(2014). 초등학생의 스마트폰 중독요인에 관한 연구. **복지행정논총**, 24(1), 47-76.
- 이건명(2018). **인공지능 튜링테스트에서 딥러닝까지**. 생능출판사.
- 이다솔, 김광민, 문상호, 권기현(2019). 인터넷·스마트폰 과의존(중독) 영향요인 분석: 의사결정나무모형을 중심으로. **정책분석평가학회보**, 29(4), 241~270.
- 이은정, 송영수, 김지하, 오수현(2020). 랜덤 포레스트를 활용한 4년제 대학 중도탈락률 예측 요인 탐색: 대학 수준 결정요인을 중심으로. **교육공학연구**, 36(1), 191 - 219.
- 이정은, 김다솜, 조일현(2020). 동영상 기반 학습 환경에서 머신러닝을 활용한 행동로그의 학업성취 예측 모형 탐색. **컴퓨터교육학회논문지**, 23(2), 53-64.
- 이창묵, 이현숙, 나우열(2020). 순환신경망(RNN)을 활용한 개별 학생 학업성취도의 중단적 예측. **교육평가연구**, 33(1), 161-189.

- 이충효(2017). 초·중·고등학생들의 스마트폰 중독에 영향을 미치는 요인에 대한 메타분석. **지역복지정책**, 28, 130~157.
- 정보통신정책연구원(2019). **어린이와 청소년의 휴대폰 보유 및 이용행태 분석**. KISDI STAT Report Vol 19-18.
- 조선희, 전경숙(2016). 대학생의 스마트폰 중독 관련 요인. **대한스트레스학회**, 24(1), 13-22.
- 조현국(2018a). 과학의 미적 가치 인식에 대한 학습자의 응답에 대한 네트워크 및 감성 분석과 머신 러닝을 활용한 탐색적 예측 평가. **학습자중심교과교육연구**, 18(20), 1325-1346.
- _____ (2018b). 머신 러닝을 활용한 이러닝 학습 환경에서의 학습자 성취 예측 모형 탐색. **학습자중심교과교육연구**, 18(21), 553-572.
- 주석진, 조성심(2015). 청소년의 스마트폰 중독 판별요인에 관한 연구: 일평균 스마트폰 사용시간, 충동성, 교우관계, 모의 권위주의형 부모양육태도, 학교부적응을 중심으로. **청소년복지연구** 17(1), 97-118.
- 하형석, 김성은, 이용혜 (2019). **한국 아동·청소년 패널조사 X : 사업보고서**. 한국청소년정책연구원.
- 한국정보화진흥원(2011). **스마트폰 중독 진단척도 개발 연구**.
- Buctot, D. B., Kim, N., & Kim, J. J. (2020). Factors associated with smartphone addiction prevalence and its predictive capacity for health-related quality of life among Filipino adolescents. *Children and youth services review*, 110.
- Chou, H. L., & Chou, C. (2019). A quantitative analysis of factors related to Taiwan teenagers' smartphone addiction tendency using a random sample of parent-child dyads. *Computers in human behavior*, 99, 335-344.
- Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York, NY: Springer New York.

ABSTRACT

Performance Analysis of Machine Learning Algorithms Using Data related to Smartphone Addiction of Elementary School Students

Lee, Chul-hyun (Gyeongin National University of Education)

The purpose of this study is to explore the optimal machine learning algorithm by comparing and analyzing the performance of machine learning algorithms using data related to smartphone addiction of elementary school students. For this purpose, data including factors affecting smartphone addiction and propensity for smartphone addiction were collected from KCYPS 2018 data based on prior research. The performance of the machine learning model calculated by applying the collected data to machine learning algorithms K-Nearest Neighbor, Logistic Regression, Support Vector Machines, Decision Tree, and Random Forest was evaluated. As a result, the relative importance of predictors was high in order of smartphone usage time, academic enthusiasm, attention, self-esteem, depression, aggression, and game usage time. The performance of each machine learning algorithm was similar, of which random forest had the best performance. The implications of the study were as follows: First, the models produced in this study will guide the direction of Practical Arts classes on cyber addiction prevention properly, and trigger research on artificial intelligence application in other areas of Practical Arts. Second, the significance of the research can be found in that it has confirmed for the first time in Korea the possibility of applying artificial intelligence technology to smartphone addiction research of elementary school students. This could contribute to the exploration of a longitudinal prediction model of smartphone addiction tendency for elementary school students. Third, the relative importance of factors affecting smartphone addiction shown in the study will help elementary school students find directions for educational prescriptions to prevent smartphone addiction.

Key Words : artificial intelligence, elementary school students, machine learning, prediction model, smartphone addiction