

기계학습을 활용한 중학교 2학년 학생 우울 예측 변수 탐색: KCYPs 2018 제5차 자료 분석*

김형관** · 유진은***

I 알기 쉬운 개요

우울은 불안이나 수면장애 등의 다양한 문제와 관련되는 정신건강 위험요인입니다. 본 연구는 청소년의 우울을 예방할 목적으로 한국아동·청소년패널자료를 두 가지 기법으로 분석하고, 두 기법 중 예측 성능이 좋으면서 결과 해석도 용이한 기법을 최종 기법으로 선택하였습니다. 연구 결과, 수백 개의 개인 및 환경 변수 중 다음과 같은 주요 우울 예측변수를 파악하였습니다. 자아존중감, 정서문제, 신체증상, 그릿, 사이버 비행 경험과 같은 예측변수는 선행연구에서도 다루었던 변수로, 예측 방향성 또한 선행연구와 일치하였습니다. 학업 무기력과 협동심 등의 변수는 예측 방향성이 선행연구와 달랐습니다. 희망직업 선택 요인, 주말 TV 시청 시간, 선호하는 연예인이나 운동선수 여부, 가식적 성격, 주말에 부모님과 함께 보내는 시간 등과 같은 변수는 선행연구에서 아직 다루지 않은 새로운 예측 변수인 것으로 나타났습니다. 연구 결과를 바탕으로 중학생의 정신건강 증진을 위한 지원체계 등을 논하였습니다.

* 본 과제(결과물)는 2024년도 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 지자체-대학 협력기반 지역혁신 사업의 결과임 (2021RIS-001).

** 한국교원대학교 교육과정 및 평가 전공 박사과정,
milphy@nate.com

*** 한국교원대학교 교육학과 교수, 교신저자, jeyoo@knue.ac.kr

투 고 일 / 2024. 3. 11.
심 사 일 / 2024. 6. 11.
심사완료일 / 2024. 6. 20.

I 초록

정신건강 지표 중 우울은 청소년이 경험할 수 있는 보편적인 정서이다. 그러나 우울은 불안이나 수면장애 등 다양한 증상을 동반할 수 있는 정신건강 위험요인이기도 하다. 중학생의 우울을 관리하고 예방함으로써 청소년기 전반의 정신건강을 증진시키며 삶의 질을 개선할 수 있다. 따라서 중학생의 우울을 예측하고, 이에 기여하는 변수가 무엇인지 살펴볼 필요가 있다. 본 연구의 근본 목적은 청소년의 우울에 기여하는 변인을 총체적인 관점에서 파악하는 것이다. 본 연구에서 수백 개의 개인발달과 발달환경 변수를 제공하는 대용량 자료인 한국아동·청소년패널 2018(KCYPs 2018)을 기계학습 기법(group Mnet이 random forest)으로 분석한 결과, group Mnet이 최종 연구 기법으로 선정되었으며, 총 305개 변수 중 42개가 중학생의 우울을 예측하는 변수로 선택되었다. 그중 선행연구와 회귀계수 방향이 일치하는 변수는 자아존중감, 정서문제, 신체증상, 그릿(grit), 사이버 비행 경험, 주관적 건강상태, 행복감, 친구와의 관계, 부모의 양육태도인 것으로 나타났다. 선행연구와 일치하지 않는 변수는 학업 무기력, 협동심 변수로 나타났다. 또한, 본 연구를 통해 6개의 변수가 우울의 예측에 기여하는 변수로 새롭게 발견되었다. 구체적으로 희망직업 선택 요인, 주말 TV 시청 시간, 선호하는 연예인이나 운동선수 여부는 우울을 부적으로 예측하였고, 가식적 성격, 주말에 부모님과 함께 보내는 시간, 스마트폰으로 문서 보기는 우울을 정적으로 예측하였다. 우울을 예측하는 변수로 탐색된 변수 중 기존의 선행연구와 일치하지 않거나 새롭게 발견된 변수를 고려하여 학생의 우울을 예방 및 개선할 필요가 있다. 이를 통하여 학생 정신건강 증진을 위한 지원체계와 학생의 삶의 질 개선에 실질적인 도움을 제공할 수 있을 것으로 기대한다.

주제어: 한국아동·청소년패널조사(KCYPs), 우울, Group Mnet, Random forest, 변수 탐색

I. 서 론

1. 연구의 필요성

아동·청소년의 삶의 질 파악과 정책적인 개선점이 필요하다는 사회적인 요구 증가에 따라, 지난 2022년 통계청에서는 아동·청소년 삶의 질 보고서를 발간하였다(심수진, 김은아, 2022). 아동과 청소년은 우리 사회의 보호와 돌봄이 필요한 취약집단으로, 더욱 많은 관심을 기울여야 할 필요가 있다. 아동·청소년기는 하나의 사회 구성원으로 인정하고 존중하되, 성인이 되기 이전 미래의 삶을 위한 발판을 마련하고 성장하는 데 도움이 필요한 시기이다. 그러나 해당 보고서에 따르면, 이러한 성장기의 과정에서 성인이 되지 못한 채 스스로 목숨을 끊는 사례가 시간이 지남에 따라 증가하는 추세다. 2000년 아동·청소년 인구 10만 명 당 1.2명이었던 자살률이 2021년에는 2.7명으로 2배 이상 증가하였다. 특히 중학생에 해당하는 만 12세부터 14세 청소년의 자살률 증가가 심각하다. 고등학생에 해당하는 만 15세에서 17세 청소년의 자살률이 2000년 5.6명에서 2021년 9.5명으로 2배 가량 증가하였으나, 중학생 시기의 청소년의 경우 1.1명에서 5.0명으로 5배 가까이 증가하였다. 자살생각 비율과 자살시도율에 있어서도 고등학생보다 중학생의 비율이 상대적으로 높은 것으로 나타났다. 즉, 아동·청소년 중에서도 중학생의 정신건강 문제에 더욱 관심을 기울여야 한다. 특히 정신건강 중에서도 가장 중요하게 다루어야 할 영역인 청소년의 우울(Kovacs, Obrosky, Gatsonis & Richards, 1997)에 초점을 맞춰 정신건강 문제를 살펴볼 필요가 있다.

우울은 슬프거나 불행하다는 감정을 느끼는 것으로, 일반적인 청소년이 경험할 수 있는 보편적인 정서라 하겠다(김정민, 송수지, 2014). 그러나 동시에 여러 증상을 동반하는 정신건강상 위험요인이기도 하다. Kirmayer, Robbins and Paris(1994)에 따르면, 우울로 인한 증상은 수면장애나 운동장애, 불안부터 그로 인한 신체적 증상에 이르기까지 다양하고 이질적으로 나타날 수 있다. 특히, 중학생 시기는 급격한 성장기의 불안감 등 다양한 문제를 양산하는 초기 사춘기이면서(고영희, 홍후조, 2012) 학업의 부담이 가중되는 시기로, 정신건강 측면에서의 위험요인을 관리하고 예방할 필요가 있다. 이를 위하여 중학생의 정신건강 위험요인인 우울을 예측하고, 관련 변인들이 무엇인지 파악하는 것이 중요하다.

중학생의 우울과 여러 변인과의 관계를 밝히는 연구는 다수 찾아볼 수 있다. 선행연구에서는 개인적인 측면의 자아존중감이나 학업무기력, 환경적인 측면의 부모 또는 친구와의 관계 등이 우울을 설명하는 대표적인 변인이다. 그러나 기존 연구는 연구자가 선택한 소수의 변수로 설명적 모형을 제시하는 것에 그친 것으로, 수많은 변수 가운데 어떠한 변수가 우울과

관련되어 있는지 탐색한 것은 아니다. 대용량 자료에 기계학습(machine learning) 기법을 적용할 경우, 수백 개에 이르는 변수를 하나의 모형에 투입함으로써 선택된 변수만이 아닌 총체적인 관점에서 예측 변수를 탐색할 수 있다. 본 연구에서는 기계학습 기법으로 가능한 모든 변수를 하나의 모형에서 고려하여 우울 예측에 기여하는 변수를 탐색하기 위하여 한국아동·청소년패널조사 2018(이하 KCYPS 2018)을 활용하였다.

2. 연구목적 및 연구문제

본 연구의 주목적은 중학생의 정신건강 위험요인인 우울을 예측하는 것으로, 청소년기 우울의 원인인 개인의 자존감이나 학업 등의 개인적인 요소와 부모나 친구와의 관계 등 환경적인 측면을 다각적으로 살펴보는 것이 중요하다. 기존의 연구에서는 전통적인 통계모형을 사용하여 수많은 변수를 하나의 모형안에서 다루지 못하기 때문에, 연구자의 판단으로 선택된 변수만이 사용된다. 그러나 기계학습 기법을 적용하면 자료의 가능한 모든 변수를 고려하여 중학생의 우울을 예측할 수 있다는 차별점이 있다. 기계학습 기법은 현재 자료에 대한 설명(explanation)보다는 기존 자료를 통해 구축된 모형을 통한 새로운 자료에 대한 예측(prediction)에 그 초점이 있어(유진은, 2021), 본 연구의 목적에 부합하는 방법이다. 기계학습 기법 중 사회과학 연구에서 가장 널리 사용되는 지도학습 기법으로는 별점회귀모형과 랜덤 포레스트(random forest)가 대표적이다. 본 연구에서는 별점회귀모형 기법 중 group Mnet과 랜덤 포레스트를 사용하여 모형 예측력을 비교하고, 최종 연구 모형에서 중학생의 우울 예측에 기여하는 변수가 무엇인지 파악하였다. 연구문제를 정리하면 다음과 같다.

첫째, KCYPS 2018 중학생의 우울 분석 시 group Mnet과 랜덤 포레스트 중 예측력이 더 높은 기계학습 기법은 무엇인가?

둘째, KCYPS 2018에서 중학생의 우울 예측에 기여하는 변수는 무엇인가?

II. 선행연구 분석

본 연구는 기계학습 기법을 적용하여 KCYPS 2018에 포함된 수백 개의 가능한 모든 변수를 총체적으로 탐색하는 예측 연구이다. 기계학습 기법으로 우울 예측 변수를 탐색한 선행연구가 없었기 때문에 전통적인 기법으로 통계적 유의성 검정을 적용한 연구를 살펴보았다. 우울 선행연구 중 본 연구와 같은 자료를 사용한 연구를 분석하여 중학생의 우울과 학생

특성 및 배경과의 관계를 살펴보았다. ‘우울’을 측정하는 척도가 매우 다양하므로 연구 간 결과 비교가 쉽지 않은 상황에서, 같은 척도로 측정한 대용량 자료인 KCYPS 연구로 국한한 선행연구와의 비교가 의미가 있다고 하겠다. 우울을 반응변수로 사용하여 변인의 직접적인 영향을 살펴본 연구와 더불어 매개효과와 같은 직간접적 영향을 살펴본 연구를 포함한다. KCYPS 2018의 학생 패널자료는 크게 개인발달과 발달환경의 두 대영역으로 구성되어, 선행 연구에서 사용된 변수를 기준으로 개인발달 영역과 발달환경 영역으로 분류하여 제시하였다.

1. 개인발달 영역

자아존중감, 학업 무기력, 그릇은 우울 선행연구에서 가장 빈번하게 사용된 변수이다. 구체적으로 자아존중감과 우울이 서로 부적으로 유의한 관계가 있었고(김호성, 황정하, 2022; 김도희, 2022; 한은지, 하문선, 2021; 전숙영, 2021), 우울은 학업 무기력에 정적인 영향을 미치는 것으로 나타났다(김은예, 한수연, 박용한, 2023; 김호성, 황정하; 2022). 한편, 그릇은 우울과 서로 부적으로 유의한 관계가 있었다(남수현, 배은정, 2023; 김선미, 김유선, 류도희, 김영희, 2022; 김나현, 안상근, 2022; 전숙영, 2021).

우울과 정서문제 간 관계를 살펴본 연구도 많았다. 이를테면 정서문제 중 사회적인 위축과 주의집중 문제가 우울에 정적으로 유의한 영향을 미쳤고(김수연, 안도희, 2021; 박이슬, 김선미, 김영희, 2022), 우울은 공격성에 정적으로 유의한 영향을 미치는 것으로 나타났다(염효원, 오현숙, 2024; 김두경, 송지희, 김윤주, 2022; 하문선, 김은선, 한은지, 안수연, 2021). 또한, 우울로 인해 발현되는 신체화 증상(예: 복통, 두통)에 초점을 맞춘 연구에서는 우울이 신체화 증상에 정적으로 유의한 영향을 미쳤다(남수현, 배은정, 2023; 이선영, 2022). 이외에 사이버 비행 관련 연구(염효원, 오현숙, 2024; 김두경 외, 2022; 하문선 외, 2021), 협동심 관련 연구(손인봉, 최정아, 2021), 주관적 건강상태와 행복감 관련 연구(채인석, 장효열, 2021)도 있었다. 사이버 비행은 우울이 정적으로 영향을 미쳤고, 협동심은 우울이 부적으로 유의한 영향을 미쳤다. 주관적 건강상태와 행복감의 관계에서 우울은 두 변수를 부적으로 부분 매개하는 것으로 나타났다.

2. 발달환경 영역

발달환경 영역은 크게 친구와의 관계와 부모와의 관계에 대한 구인으로 나뉘 살펴볼 수 있다. 긍정적인 친구와의 관계는 우울과 부적으로 유의하였다(김은예 외, 2023; 손인봉, 최정

아, 2021). 부모와의 관계 관련 연구는 학생이 자각하는 부모의 양육태도 관련 연구가 대부분이었다. 부모의 양육태도는 부정적 양육태도와 긍정적 양육태도로 나뉜다. 부정적 양육태도로는 거부, 강요, 비밀관성이 있으며, 긍정적 양육태도로는 따스함, 자율성 지지, 구조제공이 있다. 부정적 양육태도는 우울과 정적인 관계를 보였고(남수현, 배은정, 2023; 김나현, 안상근, 2022; 김선미 외, 2022; 한은지, 하문선, 2021; 하문선 외, 2021). 긍정적 양육태도와 우울은 대체로 부적의 관계가 있는 것으로 나타났다. 구체적으로 긍정적 양육태도의 하위 구인 중 따스함과 자율성 지지가 그러하였다(손인봉, 최정아, 2021; 전숙영, 2021). 그러나 또 다른 하위 구인인 구조제공의 경우에는 유의한 관계를 보이지 않았다는 연구도 있다(손인봉, 최정아, 2021).

III. 연구방법

1. 분석자료

KCYPS 2018은 2018부터 층화 다단계집락표집 방식으로 전국의 초4, 중1 학생을 표집하여, 7년간 추적 조사하는 종단연구 자료다. 중학생의 우울이 주된 관심사인 본 연구에서는 중1 코호트 대신, 최신 자료인 KCYPS 2018의 초4 코호트가 중학교 2학년이 되는 제5차 조사의 학생 패널자료를 활용하였다. 해당 학생 패널은 중학교 2학년에 해당하는 나이이므로 중학교 2학년 학생으로 간주하였다.

1) 데이터 클리닝

중학교 2학년 학생 총 2,311명이 응답하여, 437개의 변수를 포함하는 자료에 대하여 데이터 클리닝 이후 반응변수를 제외한 설명변수는 총 305개로 정리되었다. 데이터 클리닝 과정은 다음과 같다. 우선, 개인 및 가구 ID, 가중치, 개방형 문항 변수, 형제자매 패널용 문항, 생년월일 등을 삭제하였다. 다음으로 시간과 관련하여 기상, 취침(YTIM1A01, YTIM1B01 등), 몇 번째 형제자매와 얼마만큼의 시간을 보냈는지 묻는 변수(YFAM1B01 YFAM1B02, 등)는 연구에서 불필요하다고 판단하여 삭제하였다. 부모님과 함께 보내는 시간(YFAM1A01, YFAM1A02 등)은 필요하다고 판단하여, 시간과 분으로 나뉜 변수를 시간 단위로 통합하였다. 다음으로, 결측비율이 30% 이상인 문항 22개(YACT1B01, YFUR4A00 등)를 삭제하였다. 마지막으로 분산이 0에 수렴하는 41개 변수(YTWIN, YTIM1G02 등)를 삭제하였다. 해당

변수는 문항에 대한 학생 응답이 거의 동일하여 분석의 의미가 없는 변수이기 때문이다.

2) 반응변수

본 연구의 반응변수는 중학생의 우울로, KCYPS 2018의 10개의 문항으로 구성된 우울과 관련한 변수의 평균을 사용하였다. 10개 문항의 평균은 1.754이고 표준편차는 .530이며, 중앙값은 1.3으로 최소값은 1, 최대값은 4의 범위이다. 해당 10개 문항의 신뢰도 분석을 진행하였으며, 문항내적일치도인 크론바흐의 알파값은 .894로 높은 신뢰도를 보였다. 따라서, 10개 문항의 평균을 연속형 반응변수로 사용하는 것에 무리가 없다고 판단하였다.

3) 결측치 대체 및 더미변수 생성

자료 정리 이후 모든 문항에 대한 결측을 보면 총 305개의 설명 변수 중, 결측이 있는 변수는 29개로 나타났다. 다만, 해당 변수 중에서 27개의 변수는 스마트폰 사용 여부에 대한 사용 목적을 묻는 문항으로, 스마트폰을 사용하지 않는다고 답한 학생은 구조적으로 결측일 수밖에 없는 문항이다. 따라서, 해당하는 27개 변수의 결측치인 23명에 대해서는 0으로 대체하여 스마트폰을 사용하지 않는다는 것으로 나타내었다. 나머지 결측이 있는 2개 문항(ARA1A, ARA2A)은 4개의 결측이 있는 것으로 나타났다. 결측치를 처리하기 위하여 k-NN(k-Nearest Neighbors; k-최근접 이웃) 결측 대체를 사용하였다. k-NN 결측 대체는 결측치를 중심으로 k개의 자료에 대한 거리로 예측하여 결측을 대체하는 알고리즘이다. 여기에서 k값은 사용자가 설정할 수 있는데, 본 연구에서는 k값으로 $\sqrt{N}$ (N: 전체 사례수)을 내림한 값인 48을 사용하였다(유진은, 2021). 결측치 대체 이후, 범주형 변수에 대하여 더미변수를 생성하였다. 본 분석자료의 범주형 변수는 2개(YFUR1A00, YFAM3B00)로 각 5개, 6개 수준인 변수다. 따라서, 최종적으로 설명변수는 더미변수를 포함하여 총 312개로 정리되었다.

2. 분석 방법

1) Group Mnet

group Mnet(이하 Mnet)은 기계학습 기법의 하나인 별점회귀모형이다. 별점회귀모형은 선형모형인 기존 회귀모형 식에 별점을 부과하여 예측에 대한 영향력이 낮은 변수의 계수를

축소한다. 벌점이 부과되지 않은 회귀모형의 계수는 불편향 추정치 중 가장 작은 MSE(mean squared error)를 갖는 반면, 벌점회귀모형은 편향된 추정치인 대신 분산을 줄여 MSE 자체가 작아지게 된다(Hastie, Tibshirani & Friedman, 2009). 본 연구의 주 연구방법인 Mnet은 벌점회귀모형 중 하나인 MCP(Minimax Concave Penalty)에 ridge(능형회귀) 벌점항을 목적함수에 추가한 형태로, 다중공선성 문제를 처리할 수 있다는 장점이 있다. 또한 MCP의 목적함수는 회귀계수가 작은 경우 더 큰 벌점이 부과되는 오목 벌점함수(concave penalty function)이므로, 회귀계수 추정치의 일치성을 확보할 수 있다(Huang, Breheny, Lee, Ma & Zhang, 2016; Zhang, 2010). 즉, Mnet은 영향력이 낮은 변수의 계수를 0까지 축소하여 차원 축소의 기능을 할 뿐만 아니라, 다른 모형에 비해 추정치가 안정적인 모형이라 할 수 있다. 변수 선택 시 더미코딩된 범주형 변수를 다룰 수 있는 Mnet을 group Mnet이라 한다. 선택된 변수는 반응변수 예측에 기여하는 변수로, 회귀계수를 통해 해석이 가능하다는 장점이 있다(Immekus, Jeong & Yoo, 2022; Yoo, Rho & Lee, 2022). Mnet의 하이퍼파라미터(hyperparameter)인 λ 조율(tuning)을 위하여 교차타당화(cross-validation)를 사용하였다. Mnet에 대한 더 자세한 설명은 Huang et al.(2016)을 참고하면 된다. 분석은 R의 grpreg 패키지를 활용하였다(Breheny, Zeng & Kurth, 2022).

2) 랜덤 포레스트

랜덤 포레스트(random forest; 이하 RF)는 기계학습 기법을 활용한 사회과학 연구에서 자주 사용되는 모형으로, 의사결정나무(decision tree)를 결합한 앙상블 기법이다(Breiman, 2001). RF는 부트스트랩(bootstrap)을 활용하여 다수의 의사결정나무를 생성한 후 해당 결과들을 종합하여 하나의 결과를 도출한다. RF는 대용량 자료의 변수 간 다양한 상호작용이 있는 경우 예측력이 높다는 장점이 있으나, 비선형 모형으로 해석이 쉽지 않다는 단점도 있다(유진은, 2021). 모형 해석을 위하여 변수의 중요도(importance)를 살펴볼 수 있다. 중요도는 부트스트랩 과정에서 포함되지 않은 관측치인 OOB(out-of-bag) 관측치를 통하여 계산된다. 중요도 지수는 RF를 구성하는 나무의 노드를 늘렸을 때 변수가 얼마만큼의 불순도를 감소시켰는지 볼 수 있는 Mean Decrease Gini(MDG)와 예측 오차 감소율을 고려하는 Mean Decrease Accuracy(MDA)가 있다. 두 중요도를 고려하여 연구자 판단하에 변수를 선택할 수 있다. 하이퍼파라미터의 경우 표본이 적은 경우를 제외하면 나무의 개수(ntree)는 많이 늘리더라도 200 정도에서 오차가 안정적이게 되며 크게 차이가 없다(Hastie, Tibshirani & Friedman, 2009). 설명변수 후보 개수(mtry)를 조율하기 위하여 격자 탐색(grid search) 방법으로 152~101($p/2 \sim p/3$)까지 모형 성능을 확인한 결과(Svetnik, Liaw & Tong, 2004),

MSR(mean of squared residual)은 .103~.105로 거의 차이가 없었다. 따라서 ntree와 mtry 모두 패키지 기본값인 500과 $p/3$ 을 사용하였다(Liaw & Wiener, 2022). 분석 프로그램으로는 R의 randomForest 패키지를 활용하였다.

3) 모형 평가 및 변수 선택

기계학습 기법의 예측력 평가를 위하여 자료를 분할하는 절차가 필수적이다. 일반적으로 모형을 생성하기 위한 훈련자료(train data)와 해당 모형의 예측력을 확인할 수 있도록 하는 시험자료(test data)로 원자료를 7:3의 비율로 분할한다. 본 연구에서도 훈련자료와 시험자료로 분할하고 RMSE(root mean squared error)로 예측력을 평가하였다. 분할로 인한 편향 문제를 다루기 위하여 자료 분할과 모형 적합 절차를 각 100번 시행한 후, 100개 모형의 결과를 종합하여 제시하였다.

IV. 결과

1. 분석 방법 간 모형 비교

본 연구의 목적 중 하나인 중학생의 우울을 예측하는 모형으로 별점회귀모형인 Mnet과 앙상블 기법인 RF의 예측 성능을 비교하였다. 표 1에서 각 모형의 시험자료에 대한 RMSE를 표로 정리하였다.

표 1
Mnet과 RF의 시험자료 RMSE

모형	Min	Q1	Median	Mean	Q3	Max	SD	CI	
								하한	상한
Mnet	.301	.316	.322	.323	.331	.344	.010	.304	.343
Random Forest	.299	.312	.318	.318	.323	.334	.008	.303	.333

RMSE의 평균이 Mnet과 RF에서 .323과 .318로 나타났다. RMSE 평균의 95% 신뢰구간은 Mnet이 .304에서 .343으로, RF가 .303에서 .333이었다. RMSE 평균의 95% 신뢰구간이 겹치므로 Mnet과 RF 예측 성능의 차이가 통계적으로 유의하지 않았다. 따라서, 비선형 모형으로의 한계를 지니는 RF 대신, 선형 모형으로 결과 해석이 용이한 Mnet을 최종 모형으로 선택하였다.

2. 변수 선택 결과

100개의 Mnet의 모형에서 변수별 선택된 횟수를 구간으로 나눠 종합한 결과는 표 2와 같다(Lim, Yoo, Rho & Ryu, 2022; 김형관, 유진은, 2023).

표 2

Mnet의 변수 선택횟수

	Total	≥1	≥25	≥50	≥75	=100
Mnet	305	283	84	42	24	5

총 305개의 설명변수 가운데, 1번 이상이라도 포함된 변수는 283개로 거의 대부분의 변수가 1번 이상 선택되었다. 그러나 25번 이상 선택된 변수는 약 3배 이상 줄어들어, 84개의 변수만 선택되었다. 다음으로 100개의 모형 가운데 절반 이상 선택된 변수는 42개로, 25번 이상 선택된 구간에 비하여 변수가 2배 줄어들었다. 다음으로 75번 이상 선택된 변수는 24개로, 100번 모두 선택된 변수는 5개로 나타났다. Yoo & Rho(2021)의 모의실험 연구 결과를 따라 정리한 결과, Mnet에서 50번 이상 선택된 우울 예측변수는 총 42개였다. 해당 변수들의 계수 평균과 선택 횟수 등을 개인발달 영역과 발달환경 영역의 두 표로 정리하였다(표 3과 표 4). 선택된 42개 변수의 기술통계는 <부록 1>에 제시하였다.

표 3

Mnet 변수 선택 결과(개인발달)

소영역	항목	투입변수명	변수 설명	횟수	척도	Mean	SD
자아인식	자아존중감	YPSY3A01	나는 나에게 만족한다	100	L1	-.074	.018
		YPSY3A09	나는 내가 실패자라고 느끼는 경향이 있다	99		.059	.012
		YPSY3A06	때때로 나는 내가 쓸모없는 존재로 느껴진다	93		.034	.015
		YPSY3A10	나는 나에 대해 긍정적인 태도를 지니고 있다	92		-.035	.019
		YPSY3A03	나는 내가 장점이 많다고 느낀다	75		-.030	.015
		YPSY3A02	때때로 나는 내가 어디에도 소용없는 사람이라고 생각한다	59		.022	.012
학업태도	학업 무기력	YINT2B08	공부하고 싶은 의욕이 생기지 않는다	71	L1	.026	.013
		YINT2B11	학교에서 생활하면서 행복함을 느끼지 못한다	59		.018	.013
		YINT2B01	실력 차이를 내 힘으로 극복할 수는 없다고 생각한다	58		-.018	.012

소영역	항목	투입변수명	변수 설명	횟수	척도	Mean	SD
협동의식	협동심	YPSY5A01	그룹활동을 할 때, 친구의 장점이나 능력을 적극적으로 알려준다	56	L1	.023	.014
		YPSY5A13	그룹활동을 할 때, 중요한 사항이나 문제점이 생기면 친구들과 함께 해결한다	50		.020	.014
정서문제	사회적 위축	YPSY4D04	수줍어한다	100	L1	.107	.011
		YPSY4D05	사람들 앞에 나서기를 싫어한다	100		.043	.010
		YPSY4D03	다른 사람들에게 내 의견을 분명하게 말하기 어렵다	81		.032	.015
정서문제	주의집중	YPSY4A05	주의를 기울이지 않아서 실수를 하거나 사고를 낸다	72	L1	.022	.015
정서문제	공격성	YPSY4B06	아무 이유 없이 울 때가 있다	100	L1	.092	.010
		YPSY4B05	하루 종일 화가 날 때가 있다	99		.045	.012
정서문제	신체증상	YPSY4C07	숨쉬기 어려울 때가 있다	100	L1	.073	.010
		YPSY4C06	자주 피곤하다	99		.042	.010
		YPSY4C05	입맛이 없을 때가 있다	98		.041	.011
		YPSY4C08	자주 열이 나는 것처럼 느껴진다	97		.045	.015
		YPSY4C03	속이 자주 메스거린다	96		.051	.016
		YPSY4C04	공부를 하면 배가 아플 때가 있다	88		.026	.016
		YPSY4C01	깊이 잠들지 못하고 자다가 잠에서 깨곤 한다	66		.019	.013
끈기	그릿(Grit)	YPSY7A06	나는 무엇을 완성하는데 시간이 오래 걸리면 계속 열심히 하는 것이 어렵다	95	L1	.026	.011
		YPSY7A02	나는 문제를 해결하다가 어려움이 생겼을 때 크게 좌절하지 않으며, 다른 사람들보다 빨리 좌절에서 벗어난다	67		-.014	.009
비행	사이버비행 경험 유무 및 빈도	YDLQ2A01	누군가에게 욕이나 험한 말을 직접 보낸 적이 있다	94	L2	.022	.011
건강	건강상태	YPHY1A00	건강상태 평가	84	L3	-.022	.016
행복	행복감	YPSY2A01	전반적으로 나는 ...	85	L1	-.043	.020
진로계획	희망직업 선택 요인	YFUR4B12	다양한 경험을 할 수 있다	78	L4	-.021	.012
하루일과	주말	YTIM1M02	TV를 시청하며 노는 시간	60	L5	-.007	.005
창의성	창의적 성격	YPSY6A08	가식적인	99	D1	.078	.039

L1: 1: 전혀 그렇지 않다, 2: 그렇지 않은 편이다, 3: 그런 편이다, 4: 매우 그렇다

L2: 1: 전혀 없다, 2: 1년에 1~2번, 3: 한달에 1번, 4: 한달에 2~3번, 5: 1주일에 1번, 6: 1주일에 여러 번

L3: 1: 전혀 건강하지 않다, 2: 건강하지 못한 편이다, 3: 건강한 편이다, 4: 매우 건강하다

L4: 1: 전혀 중요하지 않다, 2: 별로 중요하지 않다, 3: 중요한 편이다, 4: 매우 중요하다

L5: 1: 전혀 안함, 2: 30분 미만, 3: 30분~1시간 미만, 4: 1시간~2시간 미만, 5: 2시간~3시간 미만,

6: 3시간~4시간 미만, 7: 4시간 이상

D1: 0: 선택 안함, 1: 선택

표 4

Mnet 변수 선택 결과(발달환경)

소영역	항목	투입변수명	변수 설명	횟수	척도	Mean	SD
친구	친구관계	YEDU2A08	친구들과의 관계가 좋다	80	L1	-.027	.017
		YEDU2A12	나와 다른 아이들과는 친해질 생각이 없다	55		.016	.011
		YEDU2A13	친구들은 나의 어렵고 힘든 점에 대해 관심이 없다	54		.022	.016
양육태도	비일관성	YFAM2F04	부모님은 사전에 주의도 주지 않고 화부터 내신다	73	L1	.018	.013
	자율성 지지	YFAM2C03	부모님은 내가 중요하게 생각하는 일을 하게 해주신다	66	L1	-.026	.015
	거부	YFAM2B01	부모님이 나를 좋아하시는지 가끔 궁금할 때가 있다	62	L1	.012	.009
	따스함	YFAM2A04	부모님은 나를 특별하게 여기신다	57	L1	-.026	.013
부모관계	시간(주말)	YFAM1A2	부모님과 함께 보내는 시간(주말)	66	시간	.007	.002
스마트폰	스마트폰 사용 목적별 빈도	YMDA1B11	문서 보기	75	L6	.010	.007
팬덤활동	선호 연예인 /운동선수	YACT3A00	선호 연예인/운동선수 유무	74	D2	-.016	.011

L1: 1: 전혀 그렇지 않다, 2: 그렇지 않은 편이다, 3: 그런 편이다, 4: 매우 그렇다

L6: 1: 전혀 사용하지 않는다, 2: 거의 사용하지 않는다. 3: 가끔 사용한다, 4: 자주 사용한다

D2: 0: 없다, 1: 있다

3. 우울 예측 변수 결과 해석

1) 개인발달 영역

개인발달 영역에서는 32개의 변수가 선택되었다. 자세하게 살펴보면, 학생 스스로 실패자라고 느끼는 경향이 있거나 쓸모없는 존재로 느끼거나 때때로 소용없는 사람이라고 생각할 때 우울이 높았다(YPSY3A09, YPSY3A06, YPSY3A02). 반대로 자신에게 만족하고 긍정적인 태도를 지니며 장점이 많다고 느끼는 경우(YPSY3A01, YPSY3A10, YPSY3A03) 우울이 낮았다. 공부하고 싶은 의욕이 생기지 않거나 학교생활에 행복감을 느끼지 못하는 경우에는 우울이 높았다(YINT2B08, YINT2B11). 그러나, 학업 실력 차이를 자신의 힘으로 극복할 수 없다고 생각할수록 우울이 낮았다(YINT2B01). 그룹활동 시 친구들의 장점이거나 능력을 알려주거나 중요 사항이나 문제점을 친구들과 해결한다고 할수록 우울이 높았다(YPSY5A01,

YPSY5A13). 다음으로 수줍어하거나, 사람들 앞에서 나서기 싫어하며, 다른 사람들에게 자신의 의견을 분명히 말하기 어려워하는 학생일수록 우울이 높았다(YPSY4D04, YPSY4D05, YPSY4D03). 주의를 기울이지 않아 실수하는 학생은 우울이 높았으며(YPSY4A05), 아무 이유 없이 울 때가 있다고 답하거나(YPSY4B06) 종일 화가 날 때가 있다고 답한 학생일수록 우울이 높았다(YPSY4B05). 또한, 신체증상이 있으면 우울이 높았다(YPSY4C07, YPSY4C06, YPSY4C05, YPSY4C08, YPSY4C03, YPSY4C04, YPSY4C01). 무엇을 완성하는 데 시간이 오래 걸리면 계속 열심히 하는 것이 어렵다고 답할수록 우울이 높았고(YPSY7A06), 문제를 해결하다가 어려움이 생겼을 때 좌절하지 않고 빨리 좌절에서 벗어난다고 답할수록 우울이 낮았다(YPSY7A02). 사이버상에서 누군가에게 욕이나 험한 말을 보낸 빈도가 높을수록 우울이 높았다(YDLQ2A01). 건강상태(YPHY1A00)와 행복감(YPSY2A01)은 지표가 높을수록 우울이 낮았다. 다음으로 희망직업 선택 요인으로 다양한 경험을 중요하게 생각할수록 우울이 낮았다. 마지막으로 주말에 TV를 시청하며 노는 시간이 높을수록 우울이 낮았고(YTIM1M02), 스스로 가식적이라고 생각하는 학생은 우울이 높았다(YPSY6A08).

2) 발달환경 영역

발달환경 영역에서는 친구관계, 양육태도 중 비밀관성 자율성지지, 거부, 따스함, 부모관계에서의 주말에 함께 보내는 시간, 스마트폰 사용의 목적으로 문서 보기, 팬덤활동 선호 연예인이나 운동선수 유무 여부에 해당하는 변수 10개가 선택되었다. 해당 변수들은 친구나 부모와의 관계와 관련된 변수가 대부분이었고, 스마트폰과 팬덤활동과 관련한 문항만이 하나씩 선택되었다. 우선 친구관계에서는 친구들과의 관계가 좋다고 답할수록 우울이 낮았고(YEDU2A08), 다른 아이들이 자신과 친해질 생각이 없다고 생각하거나 친구들이 자신의 힘든 점에 관심이 없다고 생각할수록 우울이 높았다(YEDU2A12, YEDU2A13). 다음으로 부모의 양육태도에 대한 학생 설문에서는 비밀관성인 사전에 주의도 주지 않고 부모님이 화부터 낸다고 생각하는 학생의 경우 우울이 높았고(YFAM2F04), 자율성 지지인 학생이 중요하게 생각하는 일을 하게 해준다고 답할수록 우울이 낮았다(YFAM2C03). 거부에 해당하는 변수인 학생이 부모가 자신을 좋아하는지 의아한 경우가 있다고 답할수록 우울이 높았으며(YFAM2B01), 부모가 자신을 특별하게 여긴다고 생각할 경우 우울이 낮았다(YFAM2A04). 부모와의 관계에 있어서 주말에 부모와 함께 보내는 시간이 많으면 우울이 높게 나타났다(YFAM1A2). 친구관계와 부모관계를 제외하고 발달환경 영역에서는 스마트폰과 팬덤활동 관련 변수가 선택되었다. 스마트폰 사용의 목적으로 문서 보기의 빈도가 높은 경우 중학생의 우울이 높았고(YMDA1B11), 선호하는 연예인이나 운동선수가 있는 경우 우울이 낮았다(YACT3A00).

V. 결론 및 제언

1. 결론

본 연구는 기계학습 기법을 적용하여 비교를 통해 우수한 모형을 찾고, 수백 개의 변수 중 중학생의 우울 관련 변수를 탐색한 연구로 그 의미가 있다. KCYPS 2018 우울을 예측하는 기계학습 기법을 비교한 결과, Mnet과 RF의 예측력 간 통계적으로 유의한 차이가 없어 회귀 모형을 기반으로 변수의 계수 해석이 용이한 Mnet 모형을 선택하였다(연구문제 1). 그 결과 중에서 선행연구와 일치하지 않는 변수 그리고 선행연구에서는 찾아볼 수 없는 새롭게 탐색된 변수를 정리하였다(연구문제 2). 자아존중감, 정서문제, 신체증상, 그릿, 사이버 비행, 주관적 건강상태 평가, 행복감, 친구관계와 부모의 양육태도 변수는 선행연구와 일치하였다.

1) 선행연구와 일치하지 않는 변수

우선, 학업 무기력은 선행연구에서 우울과 정적인 관계에 있었다. 본 연구에서 선택된 세 개의 변수 중 두 문항은 우울을 정적으로 예측하는 반면, ‘실력 차이를 내 힘으로 극복할 수는 없다고 생각한다’ 문항(YINT2B01)은 우울을 부적으로 예측하였다. 학업 무기력 문항 묶음의 크론바흐 알파는 .916으로 매우 높았으며, 해당 문항은 역코딩 문항도 아니었다. 그러나 해당 문항은 다른 학업 무기력 문항과 달리 자기수용의 가능성을 내포한다. 즉, 자신의 약점이나 신체 조건 등을 그대로 인식하고 인정할수록 우울은 감소하는 것으로 해석할 수 있다(정은영, 신희천, 2014).

선행연구에서 우울과 협동심은 부적인 관계인 반면(손인봉, 최정아, 2021), 본 연구에서는 협동심의 하위문항인 ‘그룹 활동 시 친구의 장점이거나 능력을 적극 알려준다’와 ‘그룹 활동 시 중요한 사항이나 문제점을 친구들과 함께 해결한다’(YPSY5A01, YPSY5A13)가 우울을 정적으로 예측하였다. 두 문항은 협동심의 하위 영역인 집단촉진과 문제해결 측정 문항으로(김태준, 오민아, 이영훈, 최경준, 2015), 그룹 활동을 원활하게 하기 위한 개인적 노력을 의미한다. 원만한 대인관계를 형성 및 유지를 위한 개인적인 노력은 스트레스를 일으키는 경험으로, 청소년 우울에 정적으로 유의한 영향을 미치며(Flynn & Rudolph, 2011), 특히 친구관계로 인한 스트레스는 우울에 정적으로 유의한 영향을 미치는 것으로 나타났다(강석임, 2022; 최예린, 이소연, 2019). 따라서, 그룹 활동에서의 원만한 관계 형성 및 유지를 위한 개인적인 노력이 스트레스를 유발하여 우울과 정적인 관계를 보인다고 해석할 수 있다.

2) 새롭게 탐색된 변수

본 연구의 우울을 예측하는 변수 탐색의 결과로 기존의 선행연구에서는 없었으나, 새롭게 발견된 변수에 대하여 정리하였다. 우선 우울을 부적으로 예측하는 변수부터 살펴보았다. 희망직업을 선택한 요인으로 다양한 경험이 중요하다고 생각할수록(YFUR4B12)이 우울을 부적으로 예측하였다. 진로와 관련하여 우울이 진로적응력(정부민, 김미애, 2023), 진로정체감(오세현, 강현아, 2018)에 미치는 영향을 살펴본 연구가 있었다. 그러나 본 연구의 변수는 희망 진로 선택에서의 개인적인 성향을 묻는 문항으로 적응력과 정체감과는 맥락이 다르다. 즉, 학생이 다양한 경험과 활동을 중요시할수록 우울이 낮을 것으로 예측할 수 있다.

다음으로 주말에 TV를 시청하며 노는 시간이 많을수록(YTIM1M02) 우울을 낮게 예측하는 것으로 나타났다. 주말에 TV를 시청하며 노는 시간은 중학생이 학업에서 벗어나 온전히 휴식을 취하는 시간으로 볼 수 있다. 이러한 점에서 주말에 TV를 보며 우울의 여러 원인 중 하나인 스트레스를 풀고 자신을 정신적으로 괴롭히던 특정한 상황이나 생각에서 벗어날 수 있다는 점에서, 우울을 낮게 예측하는데 기여하는 변수로 선택되었다고 볼 수 있다. 또한, 선호하는 연예인이나 운동선수가 있는지 여부(YACT3A00)가 우울을 부적으로 예측하는 변수로 새로이 탐색되었다. KCYPS 2010을 활용한 오세현과 강현아(2018)는 과도한 팬덤활동을 하는 학생의 경우에 우울을 의심할 수 있다고 보고하였다. 그러나 선호 연예인이나 운동선수 여부는 이와 다른 맥락으로, 외부 매체와 본보기 역할을 할 수 있는 인물에 대한 관심은 우울을 감소시킨다는 것이다. 앞서 주말 TV 시청과 관련하여, 주말 동안 TV를 시청하며 선호 연예인이나 운동선수를 감상하는 경우로 연결하여 생각할 수도 있다.

다음으로, 우울을 정적으로 예측하는 새롭게 발견된 변수를 살펴보았다. 우선 창의적 성격에서의 ‘가식적’ 문항(YPSY6A08)이 선택되었다. KCYPS 2018은 최인수와 표정민(2014)의 문항 개발 연구에서 변안한 창의적 성격 문항을 사용하였다. 해당 연구에서의 ‘가식적’ 문항은 비창의적 성향을 나타내는 문항으로, 창의성 측정에 있어서 역문항이다. 가식적(affected)이라는 의미는, 어떤 일을 제대로 하지 않거나 하는 척만 하는 등의 성격적인 특성을 의미한다. 학생 스스로 가식적인 성격이라고 생각하는 경우 우울을 높게 예측하였다. 즉, 창의적 성격 측정 문항 중 우울을 예측하는 변수로 본 문항만 선택되었다는 점에서, ‘가식적’ 문항은 개별 변수로서 학생 개인의 성격 특징으로 초점을 맞추고 이해해야 한다. 어떤 일을 가식적으로 한다고 생각하는 학생은, 본인 스스로 어떤 일을 제대로 해내고 있지 못하다고 판단할 가능성이 높다고 할 수 있다.

마지막으로 주말에 부모님과 함께 보내는 시간(YFAM1A2)과 스마트폰 사용목적으로 문서 보기(YMDA1B11)가 우울을 정적으로 예측하는 변수로 새롭게 탐색되었다. 주말에 부모와 보내는 시간이 많아질수록 우울이 높게 나타났는데, 투입된 변수 중에서 주말에 부모와의 대화 시간 변수는 선택되지 않은 점을 눈여겨볼 필요가 있다. 즉, 주말에 부모님과 대화하는 것은 우울의 예측에는 기여하지 않으며, 함께 보내는 시간이 길어질수록 우울이 높다는 것이다. 본 연구의 대상이 중학생이라는 것을 고려할 때, 주말에는 부모보다 친구와 시간을 보낼 때 우울이 낮아질 수 있는 것으로 보인다. 즉, 부모가 친구를 만나는 것을 통제한다거나 만날 친구가 없는 경우 우울이 높아질 수 있음을 함의한다. 다음으로 스마트폰으로 문서를 자주 볼수록 우울을 높게 예측하였다. KCYPS 2018의 스마트폰 사용 목적과 빈도 문항에서는 문서보기 외에도 의사소통, 게임, 동영상 시청, 학습 관련 등 다양한 문항으로 구성되어 있다. 학생들은 일반적으로 스마트폰을 소통이나 기분전환, 휴식 그리고 학습과 관련하여 사용한다(김선미 외, 2022). 그러나 스마트폰을 사용의 일반적인 목적에 해당하는 문항들은 선택되지 않고 오직 ‘문서보기’만이 선택되었다는 점에서 눈여겨 볼만하다. 다른 목적이 아닌 문서보기를 주로 하는 학생들은 소통, 휴식, 학업 등 일반적인 중학생들의 삶에서 떨어져 있음을 시사한다. 즉, 중학생이 일반적으로 관심을 가질만한 것에 대한 무관심은 우울과 정적인 관계가 있는 것으로 해석할 수 있다.

2. 제언

본 연구의 의의는 중학생의 우울을 예측하는 변수를 총체적인 관점에서 탐색하여 중학생의 우울과 관련한 변인을 면밀하게 살펴볼 수 있다는 것이다. 또한, 선행연구와 같이 하위문항으로 구성된 하나의 구인과 우울 간의 관계를 살펴보는 것이 아닌, 개별 문항과 우울과의 관계를 파악하여 예측할 수 있다는 차별점이 있다. 구인에 대한 일차원성(unidimensionality) 가정 하에 문항 묶음(item parceling)을 사용하는 경우, 기존의 통계 기법에서의 수렴 문제로부터는 자유로울 수 있으나 개별 문항의 효과는 살펴보기 어렵다(Yoo, 2018). 다시 말해, 기계학습 기법을 활용하면 개별 문항 단위로 우울과의 관계를 파악함으로써 선행연구에서 찾지 못한 새로운 변수를 탐색할 수 있다. 하나의 구인을 구성하는 개별 변수와 반응변수의 관계는 그 구인과 반응변수의 관계와 언제나 동일하지는 않기 때문이다. 그 예로, 선행연구에서 학업 무기력과 우울의 관계는 정적이었으나, 본 연구에서 학업 무기력 하위문항인 ‘실력 차이를 내 힘으로 극복할 수는 없다고 생각한다’는 부적인 관계로 나타났다. 또한 기존의 통계 기법은 설명에 초점이 맞추어져 있어 통계적 유의성 검정을 포함하여 모형을 통해 추론하여 자료를

설명하고자 하지만, 기계학습 기법은 새로운 자료가 투입되었을 때 종속변수를 예측하고자 하는데 더 중점을 두므로 두 기법 간 직접적인 비교는 어렵다는 한계가 있다.

선행연구와 결과가 일치하지 않거나 새롭게 발견된 변수에 초점을 맞춰 제언하겠다. 선행연구에서 학업 무기력은 우울을 정적으로 예측하였으나, 본 연구에서는 하위문항인 ‘실력 차이를 내 힘으로 극복할 수는 없다고 생각한다’가 학생 우울을 부적으로 예측하였다. 실력 차이를 자신의 힘으로 극복하지 못한다고 생각할수록 학업 무기력에 빠질 수는 있겠으나, 오히려 이를 인정하는 자기수용이 정신건강 및 학업의 측면에서도 도움이 될 수 있다는 점을 시사한다. 따라서, 학생이 학업 무기력에 빠지지 않도록 조심하면서도, 가정이나 또래 환경에 의한 반강제적인 과도한 목표 설정을 지양하도록 해야 한다. 즉, 자기수용을 통해 학생 자신의 상황을 인정하고, 외부에서 도움을 구할 수 있음을 자각하도록 해야 한다. 이를 위해서는 학업과 관련하여 우울을 호소하는 학생들을 판별하여 집중적인 상담을 통해서 우울을 해소할 수 있도록 도움을 줄 필요가 있다. 또한, 협동심은 선행연구에서 우울과 부적인 관계인 것으로 나타났으나, 본 연구에서 협동심의 하위문항으로 측정한 ‘그룹 활동 시 친구의 장점이나 능력을 적극 알려준다’와 ‘그룹 활동 시 중요한 사항이나 문제점을 친구들과 함께 해결한다’는 우울과 정적인 관계인 것으로 나타났다. 해당 문항은 협동심 측면보다는 원만한 그룹 활동을 위한 개인의 그룹원과의 관계에 대한 노력으로 볼 수 있고, 이는 스트레스를 유발하여, 그로 인해 우울을 보일 수 있다는 것을 눈여겨보아야 한다. 즉, 그룹 활동 자체가 대인과의 소통 및 관계를 추구한다는 점에서, 과도한 학교에서의 그룹 활동 요구는 오히려 학생의 우울이나 정신건강에는 부정적인 영향을 일으킬 수도 있다는 것을 시사한다.

다음으로 희망직업을 선택 요인 중 다양한 경험이 중요하다고 생각하는 학생일수록, 주말에 TV를 시청하며 노는 시간이 많을수록, 좋아하는 연예인이나 운동선수가 있는 경우 우울을 낮게 예측하였다. 희망직업 선택 요인과 관련하여 다양한 활동이나 경험의 긍정적인 부분을 직접 체험할 수 있도록 하는 것이 우울 경감에 중요할 수 있다. 다만, 주말에 TV를 보며 우울의 여러 원인 중 하나인 스트레스를 풀고 자신을 정신적으로 괴롭히던 특정한 상황이나 생각에서 벗어날 수 있다는 점에서, 학생에게 다양한 활동을 통한 경험도 우울을 낮추는 데 중요할 수 있으나, 자신의 상황과 잠시 떨어져 마음의 자유를 누릴 수 있는 시간을 제공하는 것도 고려해볼 수 있는 부분이다. 또한, 선호하는 연예인과 운동선수가 있는 경우 우울이 낮게 예측되었는데, 학업이나 학교생활 등에만 몰두하고 이러한 시간이 제공되지 않는 학생에게는 위와 같은 대상이 존재할 수 없기 때문이기도 하다. 이와 반대로 자신의 성격이 어떤 일을 할 때 하는 체한다고 생각하는 학생이거나, 부모와 주말에 함께 보내는 시간이 많은 학생이거나 스마트폰을 문서를 보는 데 많이 활용하는 학생일수록 우울이 높게 예측되었다. 학생의 우울을 낮추기 위해서는 학생 스스로 어떠한 활동에 있어서든 끝까지 완수할 수 있도록

록 독려해야 할 것으로 보인다. 일을 끝마침으로 인해 얻는 성취와 그에 따른 만족감을 느끼게 할 수 있다면 우울을 낮출 수 있을 것으로 기대한다. 다음으로, 청소년기의 특성상 부모와 보내는 시간보다 학생 스스로 하고 싶은 일을 찾게 하거나 또래 친구들과 상호작용을 하는 시간이 더 늘어날 수 있도록, 부모의 통제나 학업 등 외부적인 요인으로 인한 학생의 부담을 줄이는 것이 정신건강의 측면에서 더 도움이 될 것이다. 마지막으로 중학생들은 일반적으로 친구와의 소통 또는 영상 시청, 게임과 같은 소통과 휴식 그리고 유흥을 위하여 스마트폰을 주로 사용한다. 그러나 문서보기를 주로 하는 학생들은 우울이 높게 예측되었으며, 이는 중학생의 일반적인 관심사에 무관심한 것으로 사료된다. 따라서 학생이 스스로 관심 있는 주제를 가질 수 있도록 하여, 삶의 활력을 불어넣어 줄 필요가 있다.

현재에도 교육부와 교육청에서 실시되는 청소년의 정신건강 증진 지원체제로 현재 정신건강지원센터에서의 학생정서·행동특성검사를 통하여 우울과 관련하여 위험군에 속해있는 학생을 선별하거나, Wee클래스를 활용한 대면 상담이나 청소년모바일상담센터를 활용한 온라인 상담 등 학생건강 증진에 실질적 도움을 줄 수 있는 프로그램들이 운영되고 있다. 다만, 위에 해당하는 검사나 상담에 있어서 본 연구에서 나타난 기존과 다른 결과를 보이거나 새롭게 발견된 변수를 고려할 수 있다면, 학생의 우울 예방 및 개선에 도움이 되리라 생각한다. 또한 이를 통하여 정신건강의 증진과 함께 학생 삶의 질 또한 개선될 것으로 기대한다.

참고문헌

- 강석임 (2022). 고등학생의 학업 스트레스 및 친구관계 스트레스와 우울의 관계에서 자아존중감의 매개효과와 조절효과. **한국사회복지교육**, 58, 31-55. doi:10.31409/kjswe.2022.58.31
- 고영희, 홍후조 (2012). 사춘기 초기 적응 교육과정 개발을 위한 요구 분석. **한국교육학연구**, 18(2), 233-262.
- 김나현, 안상근 (2022). 청소년이 지각하는 부모양육태도, 우울, 그릿, 학업무기력의 구조적 관계: 성별에 따른 다집단분석. **아동교육**, 31(3), 19-39. doi:10.17643/KJCE.2022.31.3.02
- 김도희 (2022). 부모의 행복과 자녀의 행복의 관계에서 긍정적 양육태도, 자아존중감, 우울의 매개효과. **디지털융복합연구**, 20(3), 469-479. doi:10.14400/JDC.2022.20.3.469
- 김선미, 김유선, 류도희, 김영희 (2022). 부모의 양육태도와 청소년의 스마트폰 과의존의 관계: 그릿과 우울의 매개효과. **미래청소년학회지**, 19(1), 29-55. doi:10.34244/JFOYS.2022.19.1.2
- 김수연, 안도희 (2021). 중학생이 지각한 부모양육태도가 우울에 미치는 영향: 학업무기력과 사회적 위축의 매개효과. **한국교육문제연구**, 39(3), 177-203. doi:10.22327/kei.2021.39.3.177
- 김은예, 한수연, 박용한 (2023). 우울, 친구관계, 학업 무기력 간의 관계에서 인지적 공감능력의 조절된 매개효과. **한국청소년연구**, 34(4), 193-216. doi:10.14816/sky.2023.34.4.193
- 김정민, 송수지 (2014). 중학생의 주의집중 문제와 학교생활적응의 관계에서 우울의 매개효과. **한국청소년연구**, 25(1), 5-27. doi:10.14816/SKY.2014.25.1.5
- 김채봉, 정태영, 황성완, 김재행 (2013). 남녀 청소년의 우울 및 자살 생각과의 관련 요인. **보건의료산업학회지**, 7(2), 161-177.
- 김태준, 오민아, 이영훈, 최경준 (2015). **청소년 역량지수 측정 및 국제비교 연구II: IEA ICCS 2016** (연구보고 15-R18-3). 세종: 한국청소년정책연구원.
- 김형관, 유진은 (2023). 축구 동호회 가입 유인 및 저해 예측변수 탐색: 기계학습을 통한 2021 국민생활체육조사 자료 분석. **한국체육과학회지**, 32(6), 637-650. doi:10.35159/kjss.2023.12.32.6.637
- 김호성, 황정하 (2022). 청소년의 스마트폰 의존이 학업무기력에 미치는 영향: 우울과 자아존중감의 매개효과. **청소년학연구**, 29(12), 311-340. doi:10.21509/KJYS.2022.12.29.12.311
- 남수현, 배은정 (2023). 청소년들이 지각한 부모의 양육 태도와 신체화 증상의 관계에서 그릿과 우울의 이중매개효과. **한국학교보건학회지**, 36(1), 1-8. doi:10.15434/KSSH.2023.36.1.1

- 박이슬, 김선미, 김영희 (2022). 중학생의 스마트폰 의존이 주의집중 문제를 통해 우울에 미치는 영향: 수면의 질의 조절된 매개효과. **청소년학연구**, 29(3), 219-241. doi:10.21509/KJYS.2022.03.29.3.219
- 손인봉, 최정아 (2021). 지지적 양육행동, 우울, 협동심, 또래관계 간의 구조적 관계. **한국융합과학회지**, 10(5), 332-348. doi:10.24826/KSCS.10.5.21
- 심수진, 김은아 (2022). **아동·청소년 삶의 질 2022**. 대전: 통계청 통계개발원.
- 염효원, 오현숙 (2024). 학업 및 대인관계 긴장이 청소년의 사이버 비행에 미치는 영향: 우울과 공격성의 순차적 이중매개 효과를 중심으로. **청소년학연구**, 31(1), 179-207. doi:10.21509/KJYS.2024.1.31.1.179
- 오세현, 강현아 (2018). 청소년의 우울이 진로정체감에 미치는 영향: 팬덤 활동의 매개효과를 중심으로. **학교사회복지**, 42, 163-185. doi:10.20993/jSSW.42.8
- 유진은 (2021). **AI 시대 빅데이터 분석과 기계학습**. 서울: 학지사.
- 이선영 (2022). 스마트폰을 이용한 가족과 친구와의 소통이 청소년의 신체적 건강에 미치는 영향: 친구관계와 우울감의 매개효과. **한국청소년연구**, 33(3), 103-134. doi:10.14816/sky.2022.33.3.103
- 전숙영 (2021). 부모의 자율성 지지가 후기 아동의 그릿에 미치는 영향: 자아존중감과 우울의 직렬 이중매개효과. **한국아동심리치료학회지**, 16(4), 17-32.
- 정부민, 김미애 (2023). 청소년이 지각한 부모의 부정적 양육태도가 진로적응력에 미치는 영향: 우울과 학업무기력의 순차적 이중매개효과. **미래청소년학회지**, 20(4), 131-153. doi:10.34244/JFOYS.2023.20.4.6
- 정은영, 신희천 (2014). 수치심 경험이 우울에 미치는 영향: 자기수용의 조절효과. **상담학연구**, 15(1), 183-194. doi:10.15703/kjc.15.1.201402.183
- 채인석, 장효열 (2021). 청소년의 주관적 건강상태가 행복감에 미치는 영향: 우울의 매개효과. **한국자치행정학보**, 35(2), 257-274. doi:10.18398/kjlgas.2021.35.2.257
- 최예린, 이소연 (2019). 남녀 청소년의 공동반추와 우울증상의 관계에서 대인관계 스트레스의 조절효과. **한국청소년연구**, 30(2), 241-271. doi:10.14816/sky.2019.30.2.241
- 최인수, 표정민 (2014). 한국인의 창의성에 대한 암묵적 이론 척도 개발. **한국심리학회지: 사회및성격**, 28(1), 27-47. doi:10.21193/kjspp.2014.28.1.002
- 하문선, 김은선, 한은지, 안수연 (2021). 부모의 부정적 양육태도, 청소년의 우울, 공격성, 사이버비행 및 현실비행 간 관계. **중등교육연구**, 69(2), 209-244.
- 한은지, 하문선 (2021). 부모의 부정적 양육태도와 청소년의 자아존중감, 학업무기력 및 우울 간 관계. **학습자중심교과교육연구**, 21(24), 667-684.

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. doi:10.1023/A:1010933404324
- Breheny, P., Zeng, Y., & Kurth, R. (2022). Regularization paths for regression models with grouped covariates. Retrieved from <https://cran.r-project.org/web/packages/grpreg/grpreg.pdf>
- Flynn, M., & Rudolph, K. D. (2011). Stress generation and adolescent depression: Contribution of interpersonal stress responses. *Journal of Abnormal Child Psychology*, 39(8), 1187-1198. doi:10.1007/s10802-011-9527-1
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. New York: springer.
- Huang, J., Breheny, P., Lee, S., Ma, S., & Zhang, C. H. (2016). The Mnet method for variable selection. *Statistica Sinica*, 26(3), 903-923. doi:10.5705/ss.202014.0011
- Immekus, J. C., Jeong, T., & Yoo, J. E. (2022). Machine learning procedures for predictor variable selection for schoolwork-related anxiety: Evidence from PISA 2015 mathematics, reading, and science assessments. *Large-scale Assessments in Education*. doi:10.1186/s40536-022-00150-8
- Kirmayer, L. J., Robbins, J. M., & Paris, J. (1994). Somatoform disorders: Personality and the social matrix of somatic distress. *Journal of Abnormal Psychology*, 103(1), 125-136. doi:10.1037/0021-843X.103.1.125
- Kovacs, M., Obrosky, D. S., Gatsonis, C., & Richards, C. (1997). First-episode major depressive and dysthymic disorder in childhood: Clinical and sociodemographic factors in recovery. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(6), 777-784. doi:10.1097/00004583-199706000-00014
- Liaw, A., & Wiener, M. (2022). Breiman and Cutler's random forests for classification and regression. Retrieved from <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
- Lim, H. J., Yoo, J. E., Rho, M., & Ryu, J. J. (2022). Exploration of variables predicting sense of school belonging using the machine learning method—group Mnet. *Psychological Reports*. doi:10.1177/00332941221133005
- Svetnik, V., Liaw, A., & Tong, C. (2004). Variable selection in random forest with application to quantitative structure-activity relationship. *Proceedings of the 7th Course on Ensemble Methods for Learning Machines*.

- Yoo, J. E. (2018). TIMSS 2011 student and teacher predictors for mathematics achievement explored and identified via elastic net. *Frontiers in Psychology*, 9, 317. doi:10.3389/fpsyg.2018.00317
- Yoo, J. E., Rho, M., & Lee, Y. (2022). Onlinestudents' learning behaviors and academic success: An analysis of LMS log data from flipped classrooms via regularization. *IEEE ACCESS*, 10, 10740-10753. doi:10.1109/ACCESS.2022.3144625
- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894-942. 10.1214/09-AOS729

부록 1

Mnet에서 선택된 42개 변수의 기술통계

소영역	항목	투입변수명	최소값	최대값	평균	표준편차
자아인식	자아존중감	YPSY3A01	1	4	3.00	0.62
		YPSY3A09	1	4	1.96	0.75
		YPSY3A06	1	4	1.93	0.77
		YPSY3A10	1	4	2.95	0.66
		YPSY3A03	1	4	2.88	0.70
		YPSY3A02	1	4	2.00	0.77
학업태도	학업 무기력	YINT2B08	1	4	2.11	0.78
		YINT2B11	1	4	1.88	0.72
		YINT2B01	1	4	2.19	0.72
협동의식	협동심	YPSY5A01	1	4	2.81	0.65
		YPSY5A13	1	4	2.95	0.69
정서문제	사회적 위축	YPSY4D04	1	4	2.18	0.81
		YPSY4D05	1	4	2.18	0.88
		YPSY4D03	1	4	1.96	0.81
정서문제	주의집중	YPSY4A05	1	4	2.13	0.79
정서문제	공격성	YPSY4B06	1	4	1.75	0.73
		YPSY4B05	1	4	1.85	0.78
정서문제	신체증상	YPSY4C07	1	4	1.61	0.70
		YPSY4C06	1	4	2.16	0.89
		YPSY4C05	1	4	1.94	0.83
		YPSY4C08	1	4	1.70	0.74
		YPSY4C03	1	4	1.73	0.74
		YPSY4C04	1	4	1.75	0.74
		YPSY4C01	1	4	1.72	0.75
끈기	그릿(Grit)	YPSY7A06	1	4	2.37	0.75
		YPSY7A02	1	4	2.60	0.76
비행	사이버비행 경험 유무 및 빈도	YDLQ2A01	1	6	1.22	0.76
건강	건강상태	YPHY1A00	1	4	3.18	0.57
행복	행복감	YPSY2A01	1	4	3.11	0.52
진로계획	희망직업 선택 요인	YFUR4B12	1	4	3.08	0.76
하루일과	주말	YTIM1M02	1	7	2.98	1.49
창의성	창의적 성격	YPSY6A08	0	1	0.06	0.24
친구	친구관계	YEDU2A08	1	4	3.18	0.55
		YEDU2A12	1	4	1.79	0.73
		YEDU2A13	1	4	1.87	0.72
양육태도	비일관성	YFAM2F04	1	4	1.91	0.76
	자율성 지지	YFAM2C03	1	4	3.16	0.69
	거부	YFAM2B01	1	4	2.26	0.87
	파스함	YFAM2A04	1	4	3.14	0.69
부모관계	시간(주말)	YFAM1A2	0.00	23.83	6.38	4.56
스마트폰	스마트폰 사용 목적별 빈도	YMDA1B11	1	4	2.75	1.00
팬덤활동	선호 연예인/운동 선수	YACT3A00	0	1	0.43	0.50

ABSTRACT

Predictor exploration of depression in 8th graders via machine learning: KCYPS 2018 5th wave data analysis*

Kim, Hyeong Gwan** · Yoo, Jin Eun***

Among mental health indicators, depression is a common mood experienced by adolescents, which also poses a risk to mental health, accompanied by a variety of symptoms such as anxiety and sleep disturbances. Addressing and preventing depression in middle school students can help their quality of life by promoting mental health throughout adolescence. Thus, it is essential to investigate the variables that predict depression in this age group. Both personal and environmental factors influence depression of middle school students, which necessitates a holistic analysis with large-scale data. Particularly, the Korean Children and Youth Panel 2018 (KCYPS 2018) was utilized to predict depression in middle school students with machine learning techniques: group Mnet and random forest. After comparison of prediction performance, group Mnet was selected as the final technique, identifying 42 predictors out of 305 explanatory variables. Variables consistent with previous research included self-esteem, emotional problems, physical symptoms, grit, cyber-delinquency experience, subjective health status, the degree of happiness, relationships with friends, and parenting style. In contrast, some variables measuring academic lethargy and cooperativeness were not consistent with previous findings. This study also identified six new variables that contribute to depression prediction. Specifically, a career choice factor, time spent watching TV on weekends, and presence of favorite celebrity or athlete were negative predictors of depression, while pretentious characteristics, time spent with parents on weekends, and reading from smartphones were positive predictors. Further investigation into the newly identified variables or those inconsistent with previous studies will enhance the student mental health support system and, consequently, improve the quality of life of students.

Key Words: Korean Children and Youth Panel Survey(KCYPS), depression, group Mnet, random forest, variable selection

* This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-001).

** Korea National University of Education, Curriculum and Evaluation major, doctoral student, milphy@nate.com

*** Korea National University of Education, Department of Education, Professor, corresponding author, jeyoo@knue.ac.kr