

청소년기의 심리·정서적 요인이 대학 비진학 청년의 조기 경제활동 참여에 미치는 영향 분석

The Influence of Adolescent Psychological and Emotional Factors on Early Economic Activity Participation among Non-University Youth

김형훈¹, 박소윤², 이수현³

¹ 숭실대학교 산업정보시스템공학과 20192208

² 숭실대학교 산업정보시스템공학과 20231342

³ 숭실대학교 산업정보시스템공학과 20231359

Abstract 본 연구는 청소년기 심리 상태가 성인기 초기(만 18-19세) 진로 이행에 미치는 영향을 예측적으로 탐색하고자 하였다. 이를 위해 한국청소년패널조사(KYPS) 중 2 패널 1차년도부터 6차년도까지의 종단 데이터를 활용하였다. 연구 참여자의 청소년기 심리 상태는 다차원적인 요인으로 측정되었으며, 성인기 초기 진로 상태는 만 18-19세 시점의 상세 활동 유형으로 정의되었다.

패널 데이터의 종단적 특성과 복잡한 표본 설계를 반영하기 위해, 1차년도부터 6차년도까지 모두 응답한 패널을 대상으로 하는 6차년도 종단면 가중치를 적용한 가중치 고려 분류 분석을 수행하였다. 종속 변수인 만 18-19세 진로 상태는 원 데이터의 상세 분류 코드를 연구 목적과 데이터의 예측 가능성을 고려하여 '경제 활동 참여 상태 (취업, 창업, 가업 등)'와 '기타 상태 (대학 재학/준비, 비활동, 학업 지속 등)'의 이진 범주로 재범주화하였다. 독립 변수로는 청소년기 심리 상태 요인 점수를 활용하였다. 가중 랜덤 포레스트 모델을 훈련시키고, 별도의 평가 데이터셋에서 가중 혼동 행렬 및 가중 정확도, 가중 정밀도, 가중 재현율, 가중 F1-score 등 가중 성능 지표를 산출하여 모델의 예측 성능을 평가하였다.

분석 결과, 청소년기 심리 상태 변수는 만 18-19세 시점의 진로 상태를 '생산적 활동 참여 상태'와 '비경제활동 및 학업 관련 상태'로 구분하는 데 매우 강력한 예측력을 보였다.

본 연구는 청소년기 심리 상태가 성인기 초기 진로 선택 중 경제 활동 참여 여부를 예측하는 데 중요한 역할을 함을 실증적으로 밝혔다는 의의를 가진다. 이러한 결과는 경제 활동 비참여 고위험 청소년을 조기에 식별하고 맞춤형 진로 및 심리 지원 정책을 개발하는 데 활용될 수 있다.

주제어 : 한국아동·청소년패널조사(KCYPs), 랜덤 포레스트, 청소년기 심리 상태, 조기 경제 활동 참여

1 서론

추가 예정

2 관련 연구

추가 예정

3 데이터 소개

3.1 데이터 개요

본 연구는 청소년기 심리·정서적 요인이 성인기 초기 진로 이행에 미치는 영향을 예측적으로 분석하고자 한국청소년패널조사(KYPS)의 중2 패널 1차년도(2003년)부터 6차년도(2008년)까지의 자료를 활용하였다.

초기 1차년도 조사의 최종 확보 표본은 3,449명이었으나, 6차년도까지 모든 조사에 응답하고 본 연구의 분석에 필요한 주요 변수들에서 결측치가 없는 개체를 선별한 결과, 최종 분석 대상은 2,451명의 데이터로 구성되었다.

각 연차별 데이터는 연구 대상에 대한 5점 리커트 척도의 설문 문항 점수를 저장하고 있다. 해당 데이터는 청소년의 인적 사항, 진로(희망 직업, 진로 계획, 진로 준비 현황), 여가 및 사회 활동, 일탈 및 피해 경험, 생활 영역별 시간 배분 및 중요도, 자아관 관련 심리 문항 등 청소년의 다양한 삶의 영역에 대한 질문들로 구성되어 있으며, 연차 별로 총 560개 이상의 문항이 포함되어 있다.

3.2 요인 분석

본 연구는 한국청소년패널조사 User's Guide의 원자료 설문 문항 범주를 기반으로 요인 분석을 진행했다. 이를 통해 부모애착, 일탈적 자아 낙인, 부모 요인 스트레스, 부모 감독, 물질적 요인 스트레스, 친구 관계 스트레스, 자기신뢰감, 상급학교 의존도, 부정적 자아존중감, 학업 스트레스 등 총 10개 요인을 도출했다. 이후 각 요인의 점수를 산출하여 독립변수로 활용했다.

Table 4에는 각 연차별 요인 분석 결과가 제시되어 있다. 이는 설정된 잠재 요인 구조가 실제 관측된 데이터의 패턴과 얼마나 잘 일치하는지를 나타낸다. 이 지표들이 양호한 값을 보일수록, 가정한 심리 요인들이 실제 청소년들의 응답 패턴을 어느 정도 잘 반영하고 있다고 해석될 수 있다.

CFI(Comparative Fit Index)와 TLI(Tucker-Lewis Index)는 대부분의 연차에서 0.90 이상을 기록하며 양호한 수준을 보였다. 이는 설정된 요인 구조가 데이터의 변동을 설명하는데 있어 기준 모델 대비 비교적 적절함을 시사한다.

RMSEA(Root Mean Square Error of Approximation) 지수는 모든 연차에서 0.08 이하로 나타났다. 이 지표는 모델이 모집단 공분산 구조와 얼마나 근사적인 차이를 보이는지를 나타내며, 값이 0에 가까울수록 더 나은 적합성을 의미한다.

SRMR(Standardized Root Mean Square Residual) 지수 또한 모든 연차에서 0.08 이하로 나타났다. 이 지표는 관측된 상관과 모델에 의해 예측된 상관 간의 평균적인 표준화 차이를 나타내며, 값이 0에 가까울수록 모델의 예측이 실제와 가깝다고 해석된다.

연차별 적합도 지표 값에 다소 변동이 있었으나, 전반적으로 6년에 걸쳐 분석된 심리 요인 모델들은 데이터 패턴에 대해 비교적 일관되게 양호한 수준의 부합성을 보이는 것으로

판단된다. 이러한 적합도 결과는 본 연구에서 설정한 심리 요인 구조가 6년의 추적 기간 동안 청소년들의 심리 관련 응답 패턴을 나타내는 데 있어 어느 정도 타당한 측정 모델로서 활용될 수 있음을 지지하는 근거가 될 수 있다.

3.3 데이터 전처리

먼저, 연도별로 분리된 패널 데이터를 개인 식별자를 기준으로 횡 형태로 통합하였다. 통합된 데이터에서 본 연구에 필요한 독립 변수, 종속 변수, 그리고 가중치 변수 컬럼들을 추출하였다. 이때 가중치 변수는 초기 표본 추출 확률 및 1차년도부터 6차년도까지의 종단 탈락을 고려하여 계산된 것으로, 본 연구의 상대적으로 적은 표본으로 인한 일반화 가능성의 한계점을 보완하기 위해 사용되었다.

이후, 분석에 사용될 최종 변수들에 대해 결측치가 없는 행을 제거하여 최종 분석 데이터셋을 구축하였다. 종속 변수는 원 데이터의 상세 분류 코드를 바탕으로 2개의 범주로 그룹화하고 팩터 형태로 변환하였다. 종속 변수의 분류는 Table 5과 같다.

3.4 데이터 분포

최종 분석에 활용된 대상자의 인구사회학적 특성 및 종속 변수의 분포는 다음의 표의 내용과 같다.

Table 1: 출생년도별 분포 (조사 시점: 2008년)

출생년도	빈도	비율
88	2	0.08
89	1895	77.32
90	554	22.60

먼저, 조사 시점인 2008년을 기준으로 대상자들의 출생년도를 살펴본 결과,

Table 2: 성별 분포

성별	빈도	비율
남성	1165	47.53
여성	1286	52.47

Table 3: 종속 변수 분포

구분	빈도	비율
생산적활동참여	102	4.16
비 경제활동 및 학업 관련 상태	2349	95.84

대부분의 대상자가 1989년(77.32%) 또는 1990년(22.60%)생으로 나타났다. 이는 1차 조사 시작 시점(2003년)에 중학교 2학년이었던 표본이 6차년도 시점에 대략 만 18세(1990년생) 또는 만 19세(1989년생)에 해당함을 보여준다. 소수의 대상자(0.08%)는 만 20세로 확인되었다.

성별 분포에서는 남성(47.53%)이 1,165명, 여성(52.47%)이 1,286명으로, 전체적으로 남성보다 여성의 비율이 약간 더 높은 것으로 나타났다.

본 연구에서 최종적으로 분석에 사용된 종속 변수의 분포는 ‘생산적 활동 참여’ 범주에 해당하는 대상자가 102명(4.16%)이었으며, ‘비 경제활동 및 학업 관련 상태’ 범주에 해당하는 대상자는 2,349명(95.84%)으로 나타났다.

3.5 독립 변수 간 상관관계

최종 데이터셋에서 사용된 독립 변수 간의 상관관계는 아래 히트맵과 같은 양상을 보인다. 전반적인 상관 관계 패턴: 히트맵을 통해 파악된 전반적인 관계 패턴에 대해 서술합니다. (예: “대체로 동일 요인의 시점 간 요인 점수들은 강한 정적 상관 관계를 보였으며, 특정 다른

요인들 간에도 연차에 걸쳐 유의미한 상관 관계가 관찰되었다.”)

주요 상관 관계 강조: 제시하신 $|r| > 0.7$ 인 변수 쌍들을 포함하여, 분석적으로 중요하거나 특히 상관이 높은 몇 가지 변수 쌍을 선별하여 구체적인 상관계수 값과 함께 제시합니다. (이때는 표 형태로 제시하는 것보다 텍스트로 자연스럽게 서술하는 것이 좋습니다.)

다중공선성 언급: 이러한 상관 관계가 독립 변수들 간에 어느 정도의 다중공선성이 존재함을 시사하며, 이는 모델 해석 시 주의해야 할 부분임을 언급합니다. (하지만 사용하신 Random Forest 모델은 다중공선성에 비교적 덜 민감하다는 점도 함께 언급 가능).

Table 4: Confirmatory Factor Analysis 결과

년차	CFI	TLI	RMSEA	SRMR
1	0.929	0.918	0.048	0.042
2	0.943	0.934	0.047	0.037
3	0.937	0.927	0.048	0.043
4	0.927	0.915	0.054	0.046
5	0.935	0.924	0.052	0.045
6	0.906	0.891	0.064	0.059

Table 5: 종속 변수 그룹

종속 변수	원 데이터 값
생산적활동참여	고등학교는 졸업하였고 현재 취업(임시직, 일용직포함)해있는 상태이다 고등학교는 졸업하였고 현재 창업해있는 상태이다 고등학교는 졸업하였고 현재 무보수로 가업을 돕고 있다 현재 취업(임시직, 일용직포함) 상태에서 대학 생활(방통대, 야간대 등)도 병행하고 있다 현재 창업 상태에서 대학 생활(방통대, 야간대 등)도 병행하고 있다 현재 무보수로 가업을 돕고 있는 상태에서 대학 생활(방통대, 야간대 등)도 병행하고 있다
비경제활동 및 학업 관련 상태	대학에 진학하였고 졸업할 때까지 계속 다닐 생각이다 대학에 진학하였고 조만간에 휴학할 생각이다 대학에 진학하였고 현재 휴학 중이다 대학에 진학하였고 조만간에 그만둘(중퇴할) 생각이다 대학에 진학한 후 중퇴했으며 현재 다른 대학 입학 준비 중이다 고등학교는 졸업하였고 현재 대학 입학 준비 중이다 고등학교는 졸업하였고 현재 취업(임시직, 일용직포함)을 위해 노력 중이다 고등학교는 졸업하였고 현재 창업을 준비하고 있다 대학에 진학한 후 중퇴했으며 현재 취업(임시직, 일용직포함)을 위해 노력 중이다 대학에 진학한 후 중퇴했으며 현재 창업을 준비하고 있다 고등학교는 졸업하였고 현재 가사, 육아를 돕고 있다 고등학교는 졸업하였고 현재 군 입대 대기 중이다 고등학교는 졸업하였고 현재 아무일도 하지 않고 있다 고등학교 재학 중 중학교 재학 중 고등학교, 중학교 졸업 전 중퇴

요인 점수 간 상관관계 히트맵

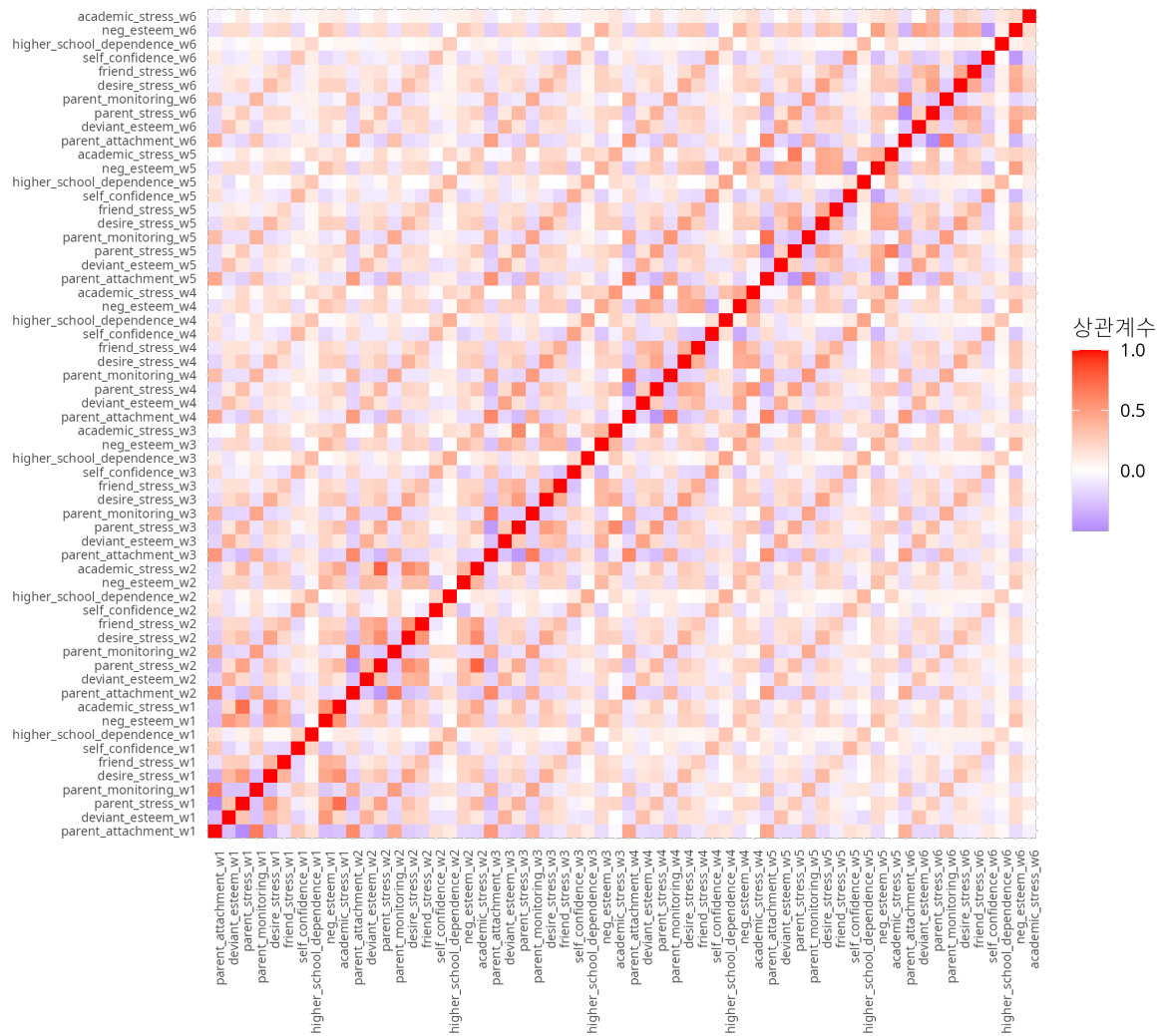


Figure 1: 독립변수 히트맵

4 연구 과정

train test set, kfold

Random forest 씬. c50, xgboost 사용.

5 연구 결과

5.1 모형별 변수 중요도 순위

5.2 모형 평가 결과

6 해석 및 논의

1. 표본 갯수 적음. 가중치 사용.
2. 종속변수의 분포가 불균형함. 이걸 어쩔 수 없음.
3. 다중공선성 문제. 하지만 랜덤 포레스트는 다중공선성에 덜 민감함.

7 참고문헌

- [한국 아동 청소년 청년 데이터 아카이브](#)