

청소년기의 심리·정서적 요인을 통한 성인 진입기 진로 안정형·탐색형 성향 분류 예측

Predicting Career Stability-Exploration Types through Psychological and Emotional Factors in Adolescence

김형훈¹, 박소윤², 이수현³

¹ 숭실대학교 산업정보시스템공학과 20192208

² 숭실대학교 산업정보시스템공학과 20231342

³ 숭실대학교 산업정보시스템공학과 20231359

Abstract 본 연구는 청소년기 심리·정서적 특성이 성인기 초기(만 18-19세) 진로 이행 패턴에 미치는 영향을 예측적으로 탐색하고자 하였다. 이를 위해 한국청소년패널조사(KYPS) 중2 패널 1차년도부터 6차년도까지의 종단 데이터를 활용하였다. 연구 참여자의 청소년기 심리·정서적 특성은 다차원적 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 6차년도 종단면 가중치를 적용한 가중치 고려 분류 분석을 수행하였다. 종속 변수인 만 18-19세 진로 상태는 Arnett(2000)의 성인 진입기 이론에 기반하여 '안정형'과 '탐색형'의 이진 범주로 재범주화하였다. 독립 변수로는 청소년기 심리·정서적 특성 요인 점수를 활용하였다. 랜덤 포레스트, XGBoost, 로지스틱 회귀분석 모델을 훈련시키고, 별도의 평가 데이터셋에서 혼동 행렬 및 정확도, 정밀도, 재현율, F1-score 등 성능 지표를 산출하여 모델의 예측 성능을 평가하였다.

분석 결과, 청소년기 심리·정서적 특성은 만 18-19세 시점의 진로 상태(안정형 vs. 탐색형) 분류에 기여하였으나, 그 예측력은 전반적으로 높지 않았다. 세 모델 모두 무작위 분류기보다는 우수한 성능을 보였으며, '안정형' 진로 상태에 대한 분류 정확도가 '탐색형'에 비해 상대적으로 양호하였다.

본 연구는 성인 진입기의 진로 이행 패턴을 청소년기 심리·정서적 발달과 연결하여 종단적으로 탐색하고, 머신러닝 기법을 활용한 진로 유형 예측 모델 구축의 초기 가능성을 제시했다는 점에서 의의를 찾을 수 있다. 본 연구의 결과는 향후 청소년 진로 상담 및 교육 정책 수립 시 고려할 수 있는 기초 자료를 제공하며, 개별 맞춤형 진로 지도 프로그램 개발을 위한 추가적인 논의와 연구 방향을 제시하는 데 활용될 수 있을 것으로 기대된다.

주제어: 한국아동·청소년패널조사(KCYPS), 랜덤 포레스트, XGBoost, Logistic Regression, Arnett 성인 진입기 이론, 청소년기 심리 상태, 진로 성향

1 서론

현대 사회에서 청소년기는 개인의 인지적·정서적 발달이 급격하게 이루어지는 결정적 시기로, 이후 성인기 삶의 다양한 영역에 중대한 영향을 미치는 토대를 마련한다. 특히 고등학교 졸업 이후 본격화되는 진로 이행 과정은 청소년기에 형성된 심리·정서적 특성과 밀접하게 연관되어 있으며, 이는 개인의 장기적인 삶의 만족도와 사회적 적응에까지 심대한 영향을 미칠 수 있다.

이러한 맥락에서 Arnett(2000)가 제안한 ‘성인 진입기(Emerging Adulthood)’ 개념은 중요한 이론적 틀을 제공한다. 만 18세에서 25세(혹은 29세) 사이로 정의되는 이 시기는, 청년들이 전통적인 의미의 성인 역할을 완전히 수행하기보다는 다양한 삶의 가능성을 적극적으로 실험하고 탐색하는 과도기적 단계이다. 이들은 완전한 책임과 구조화된 삶의 틀에 정착하기 전에 진로 방향을 수정하거나 새로운 가능성을 모색할 수 있는 ‘기회의 시기’를 경험하며, 이 과정에서의 진로 선택은 단순히 특정 교육기관으로의 진학이나 취업 여부를 넘어 개인의 심리적 안정감과 삶의 방향성 설정에 복합적인 영향을 미친다.

그러나 기존의 국내 진로 연구들은 학업 성취도, 부모의 사회경제적 지위, 학교 유형 등 외적 환경 요인에 상대적으로 집중하는 경향이 있었으며, 청소년기 동안 개인이 경험하는 내면적 심리 상태가 이후 진로 결정에 미치는 구체적인 영향과 그 기제에 대한 종단적 실증 연구는 상대적으로 부족한 실정이다.

이에 본 연구는 한국청소년패널조사(KYPS)의 중2 코호트 패널 데이터(1차년도, 중2~6차년도, 만 18-19세)를 활용하여, 청소년기에 형성된 다차원적인 심리·정서적 특성들이 성인기 초입에 나타나는 진로 이행 패턴에 어떠한 영향을 미치는지 종단적으로 분석하고자 한다. 특히, Arnett의 이론에 기반하여 청년들의 진로 상태를, 상대적으로 정착된 경로에 진입한 ‘안정형(stable)’ 집단과 아직 특정

경로에 정착하지 못하고 다양한 가능성을 모색하거나 준비 중인 ‘탐색형(explorative)’ 집단으로 재범주화하였다. 이 두 집단을 분류하고 예측하기 위해 머신러닝 기반의 분석 방법(Random Forest, XGBoost, 로지스틱 회귀분석)을 적용하여 예측 모델을 구축하고 주요 예측 요인을 탐색하였다.

본 연구는 다음과 같은 세 가지 주요 측면에서 학문적 및 정책적 의의를 가진다.

첫째, 특정 시점의 횡단적 상태에 국한되지 않고 6년간의 종단 데이터를 활용하여 청소년기의 심리·정서적 특성과 성인기 초기의 진로 상태 간의 장기적인 연결성을 실증적으로 분석함으로써 기존 연구와 차별성을 가진다.

둘째, 사회적 관심과 지원이 더 집중적으로 요구될 수 있는 ‘탐색형’ 집단의 특성을 이해하고 이들의 예측 가능성을 탐색하는 것은, 향후 이들의 원활한 사회 이행을 지원하기 위한 맞춤형 진로 지원 정책 개발 논의에 기초적인 방향을 제시할 수 있다는 점에서 잠재적인 정책적 함의를 지닌다. 비록 본 연구에서 개발된 모델의 ‘탐색형’ 집단 예측 성능은 제한적이었으나, 어떤 요인들이 이 집단과 연관되는지를 파악하는 것은 향후 보다 정교한 예측 모델 개발과 효과적인 지원 전략 수립의 출발점이 될 수 있다.

셋째, 구축된 예측모델에서 도출된 주요 영향 변수들을 통해, 청소년의 진로 이행뿐만 아니라 해당 시기의 심리·정서적 안정과 발달에 영향을 미치는 다양한 요인들을 다각적으로 조명하고 이해를 심화시킬 수 있다. 특히, 진로 미결정 상태가 우울감 등 부정적 정서와 유의미한 관련이 있다는 기존 연구 결과(장슬비, 2023)를 고려할 때, 본 연구는 청소년의 진로 상태 예측을 넘어 이들의 심리·정서적 어려움에 대한 선제적 개입의 필요성을 환기하고, 보다 건강한 성인기로의 이행을 지원하기 위한 학문적·실천적 논의를 확장하는 데 기여할 것으로 기대된다.

2 관련 연구

2.1 청소년기 진로 발달과 성인기 진로 이행

진로 발달은 개인이 일생에 걸쳐 자신의 역할을 탐색하고 정립해나가는 연속적인 과정으로 이해된다. Super(1980)는 그의 전생애 발달 이론(life-span, life-space theory)에서 진로를 “한 개인의 생애 과정으로서, 개인은 일생 동안 일련의 발달 과업에 직면하고 그 과정에서 자신이 원하는 모습의 사람이 되는 방식으로 발달 과업을 수행하려 한다”고 정의하였다. 그는 특히 14세에서 24세 사이를 진로를 구체적으로 탐색하는 탐색기(exploration stage)로 명명하며, 이 시기에 이루어지는 학업 지속 여부나 직업 선택과 같은 중요한 의사결정 및 이때 형성된 진로 태도가 성인기의 직업적 삶에 지속적인 영향을 미친다고 강조했다.

Super가 제시한 탐색기의 후반부에 해당하는 청년기 초반(일반적으로 18~25세)에 대해, Arnett(2000)는 이를 전통적인 성인기 이전의 독특한 발달 단계인 ‘성인 진입기(Emerging Adulthood)’로 개념화하여 더욱 심층적인 이해의 틀을 제공했다. Arnett는 이 시기를 정체성 탐색, 불안정성, 자기 중심성, 어중간함, 그리고 가능성의 시대로 특징지으며, 불확실성과 자기 주도적 탐색이 공존하는 역동적인 전환기로 설명하였다. 실제로 이 시기의 청년들은 대학 진학, 취업, 창업, 학업 중단(휴학), 구직 활동 등 매우 다양한 경로를 경험하며, 때로는 여러 경로를 동시에 탐색하거나 경로 간 잦은 이동을 보이기도 한다. 따라서 성인 진입기의 진로 상태를 단순히 진학이나 취업 여부와 같은 이분법적 기준으로 파악하는 데에는 한계가 있으며, 다양한 이행 경로의 패턴과 그 안에 내재된 개인의 심리적 의미를 종합적으로 이해할 필요성이 제기된다. 이러한 이론적 배경에 기반하여, 본 연구에서는 성인기 초기의 다양한 진로 상태를 해당 시기의 주요 특징을 반영하는 ‘안정형’과 ‘탐색형’이라는 두 가지 주요 패턴으로 재구성하여 분석을 시도하였다.

2.2 청소년기 심리정서적 특성과 진로 발달

청소년기는 급격한 신체적, 인지적 변화와 함께 자아정체감을 확립하고 미래를 구체적으로 고민하기 시작하는 시기로, 진로 발달에 있어 매우 핵심적인 단계에 해당한다. 이 시기에 경험하는 다양한 심리·정서적 특성들은 이후 진로 목표 설정, 진로 준비 행동, 그리고 실제 진로 선택에 중요한 영향을 미치는 것으로 알려져 있다. 특히, 진로 성숙도와 자기효능감은 청소년의 성공적인 진로 발달을 예측하는 핵심적인 내적 변수로 간주된다.

2.2.1 진로 자기효능감

자기효능감은 특정 과제를 성공적으로 수행할 수 있다는 개인의 신념을 의미하며(Bandura, 1997), 이는 다양한 영역에서의 성취 행동을 예측하는 중요한 심리적 자원이다. 특히 진로 발달 맥락에서는 진로 자기효능감(career decision self-efficacy), 즉 진로와 관련된 의사결정이나 과제를 성공적으로 수행할 수 있다는 자신감이 중요한 역할을 한다. 진로 자기효능감이 높은 청소년은 적극적으로 진로 정보를 탐색하고, 명확한 진로 목표를 설정하며, 어려움에 직면했을 때에도 쉽게 포기하지 않고 문제를 해결하려는 노력을 보인다.

이처럼 청소년기에 형성되는 긍정적인 자아 인식, 스트레스 대처 능력, 부모 및 또래와의 관계 등 다양한 심리·정서적 요인들은 진로 성숙도와 자기효능감의 발달에 기여하며, 이는 궁극적으로 합리적이고 만족스러운 진로 선택으로 이어질 수 있다. 즉, 청소년의 진로 발달은 인지적 준비나 사회적 환경뿐만 아니라, 내면의 심리적 안정성과 건강함에 큰 영향을 받는다고 할 수 있다.

그러나 청소년기의 심리·정서적 특성과 실제 성인기 초기 진로 이행 간의 관계를 탐구한 기존 연구들은 몇 가지 한계를 지닌다. 다수의 연구가 특정 시점의 횡단적 설문조사에 의존하고 있으며, 연구 대상 또한 대학생 집단에 집중되는 경향을 보여왔다. 이로 인해 진로 불안, 성숙도, 만족도

등의 주요 지표들이 주로 대학생 시기의 경험을 바탕으로 측정되었으며, 고등학교 졸업 직후 또는 보다 광범위한 성인기 초기 청년들의 다양한 진로 경로와의 장기적인 연결성을 검증한 연구는 상대적으로 부족하다. 특히, 청소년기에 경험한 구체적인 심리·정서적 특성들이 시간의 흐름에 따라 어떻게 변화하고, 이러한 변화가 이후 실제 진로 이행 결과와 어떠한 관계를 맺는지 장기적인 패널 데이터를 통해 심층적으로 검토한 연구는 매우 드문 실정이다.

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 2,437명의 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 데이터 전처리

3.2.1 요인 분석

먼저 한국청소년패널조사(KYPS)의 중2 패널 1차년도부터 5차년도까지 각 연도별 데이터에 대해 확인적 요인 분석(Confirmatory Factor Analysis, CFA)을 시행하여 다수의 설문 문항들을 소수의 잠재 요인으로 축약하였다. 요인 구조는 KYPS User's Guide의 원자료 설문 문항 범주를 기반으로 구성하였으며, 각 차수별 CFA 시 해당 연도의 횡단 가중치를 표본 가중치로 반영하여 분석의 대표성을 높이고자 하였다. 이를 통해 각 차수별로 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 요인이 도출되었다. 해당 요인들과 관련된 자세한 문항은 Table 4에서 확인할 수 있다.

이후, 각 개인의 차수별 요인 점수를 산출하였고, 이 점수들을 개인 식별자를 기준으로 병합하여 횡단면 형태의 독립변수 데이터셋을 생성하였다. 분석의 명료성을 위해, 종속변수가 측정된 6차년도의 문항들은 독립변수에서 제외하였다.

Table 5에는 각 연차별 확인적 요인 분석 결과 산출된 모델 적합도 지수가 제시되어 있다. 이 지표들은 연구자가 설정한 잠재 요인 구조가 실제 관측된 데이터의 패턴과 얼마나 잘 부합하는지를 평가하는 기준으로, 양호한 값을 보일수록 가정한 심리 요인들이 실제 청소년들의 응답 패턴을 타당하게 반영하고 있다고 해석할 수 있다.

CFI(Comparative Fit Index)와 TLI(Tucker-Lewis Index)는 모든 연차에서 일반적으로 양호한 적합도의 기준으로 간주되는 0.90 이상을 기록하여, 설정된 요인 구조가 데이터를 비교적 잘 설명함을 시사한다.

RMSEA(Root Mean Square Error of Approximation) 지수는 모든 연차에서 우수한 적합도의 기준인 0.04 이하로 나타났다. 이

지표는 모델이 모집단 공분산 구조와 얼마나 근사적인지를 나타내며, 값이 낮을수록 더 나은 적합성을 의미한다.

SRMR(Standardized Root Mean Square Residual) 지수 또한 모든 연차에서 우수한 적합도 수준인 0.04 이하로 나타났다. 이 지표는 관측된 상관과 모델에 의해 예측된 상관 간의 평균적인 표준화 차이를 나타내며, 역시 값이 낮을수록 모델의 예측이 실제 데이터와 가깝다고 해석된다.

종합적으로, 5개년에 걸쳐 각 차수별로 분석된 심리 요인 모델들은 데이터 패턴에 대해 일관되게 양호한 수준의 적합도를 보이는 것으로 판단된다. 이러한 결과는 본 연구에서 사용된 10개의 심리 요인 구조가 5년의 추적 기간 동안 청소년들의 심리 관련 응답 패턴을 측정하는 데 있어 타당한 측정 모델로서 활용될 수 있음을 지지한다.

3.2.2 종속 변수

종속변수는 6차년도의 진로 상태 관련 문항 응답을 바탕으로 아래 Table 6에 명시된 바와 같이 '안정형(stable)'과 '탐색형(explorative)'의 두 범주로 재분류되었다.

3.3 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	1884	77.3
90	551	22.60

Table 2: 성별 분포

성별	빈도	비율
남성	1157	47.47
여성	1280	52.52

Table 3: 종속 변수 분포

구분	빈도	비율
안정형	1369	56.17
탐색형	1068	43.82

먼저, 6년차 조사 시점인 2008년을 기준으로 대상자들의 출생년도를 살펴본 결과, 대부분의 대상자가 1989년(77.3%) 또는 1990년(22.6%)생으로 나타났다. 이는 1차 조사 시작 시점(2003년)에 중학교 2학년이었던 표본이 6차년도 시점에 대략 만 18세 또는 만 19세에 해당함을 보여준다. 소수의 대상자(0.08%)는 만 20세로 확인되었다.

성별 분포에서는 남성(47.47%)이 1,157명, 여성(52.52%)이 1,280명으로, 전체적으로 남성보다 여성의 비율이 약간 더 높은 것으로 나타났다.

본 연구에서 최종적으로 분석에 사용된 종속 변수의 분포는 '안정형' 범주에 해당하는 대상자가 1369명(56.17%)이었으며, '탐색형' 범주에 해당하는 대상자는 1068명(43.82%)으로 나타났다.

6년차 조사 시점 기준 연구 대상자의 지역 분포는 [그라프](#)에 제시된 바와 같다. 전체적으로 서울과 경기 지역에 표본이 치중되는 분포를 보이고 있다.

3.4 독립 변수 간 상관관계

본 연구의 최종 분석 데이터셋에 포함된 독립변수들 간의 상관관계를 살펴본 결과, 몇 가지 주목할 만한 패턴이 확인되었다. 변수

간 상관관계수 중 절대값이 높은 상위 10개 조합은 아래 [상관관계 표](#)에 제시되어 있으며, 전체적인 상관관계 패턴은 상관관계 히트맵을 통해 시각적으로 파악할 수 있다. 이러한 패턴들은 여러 연차에 걸쳐 비교적 일관된 양상을 보였으며, 동일 요인의 경우 인접한 연차 간에 특히 높은 상관관계를 나타내는 경향이 관찰되었다. 이는 청소년기 심리사회적 특성의 발달적 연속성을 시사한다.

접근이 필요함을 시사한다.

가장 두드러지는 패턴 중 하나는 부모감독 요인과 부모애착 요인 간의 관계이다. 본 연구에서 부모감독은 “청소년이 외출했을 때 부모가 자녀의 소재, 동반자, 활동 내용, 귀가 시간 등을 얼마나 잘 알고 있는지에 대한 인지 수준”을 묻는 문항들을 통해 측정되었다. 이렇게 측정된 부모감독 요인은 모든 조사 연차에서 부모애착 요인과 강한 정적 상관관계를 일관되게 나타냈다. 이는 청소년이 부모와 긍정적이고 안정적인 애착 관계를 형성하고 있을수록, 부모 또한 자녀의 일상적인 활동에 대해 더 많은 정보를 인지하고 있을 가능성이 높음을 시사한다. 즉, 친밀한 부모-자녀 관계가 부모의 자녀에 대한 이해도를 높이는 기반이 될 수 있음을 보여준다.

다음으로, 부모애착과 다양한 스트레스 요인들 간에는 전반적으로 부적 상관관계 패턴이 확인되었다. 구체적으로, 부모애착 점수가 높을수록 부모 스트레스, 친구 스트레스, 학업 스트레스 등의 점수는 낮은 경향을 보였다. 이는 청소년이 부모와 안정적인 애착 관계를 형성하는 것이 다양한 생활 영역에서 경험하는 스트레스를 완화하는 보호 요인으로 작용할 수 있음을 시사한다.

또한, 각종 스트레스 요인들 간에도 유의미한 정적 관련성이 다수 관찰되었다. 특히, ‘부모 스트레스’와 ‘학업 스트레스’ 간에는 모든 연차에서 상대적으로 강한 정적 상관관계가 일관되게 나타났다. 이는 특정 영역에서 경험하는 스트레스가 독립적으로 존재하기보다는 다른 영역의 스트레스 경험과 밀접하게 연동되어 있을 가능성을 보여주며, 청소년의 스트레스 문제를 다룰 때 다각적인

Table 4: 심리 요인별 변수명 및 설문 문항

요인	변수명	설문 문항
부모애착	parent_attachment	부모님과 나는 많은 시간을 함께 보내려고 노력하는 편이다 부모님은 나에게 늘 사랑과 애정을 보이신다 부모님과 나는 서로를 잘 이해하는 편이다 부모님과 나는 무엇이든 허물없이 이야기하는 편이다 나는 내 생각이나 밖에서 있었던 일들을 부모님께 자주 이야기하는 편이다 부모님과 나는 대화를 자주 나누는 편이다
일탈적 자아 낙인	deviant_esteem	나는 나 자신이 문제아라고 생각한다 나는 나 자신이 문제아라고 생각한다 주위 사람들은 나를 문제아라고 생각한다 주위 사람들은 나를 비행청소년이라고 생각한다
부모에 의한 스트레스	parent_stress	성적 때문에 부모님으로부터 스트레스를 받는다 부모님과 의견 충돌이 있어서 스트레스를 받는다 부모님의 지나친 간섭으로 스트레스를 받는다 부모님과 대화가 안 통해서 스트레스를 받는다
부모감독	parent_monitoring	내가 외출했을 때 부모님은 내가 어디에 있는지 대부분 알고 계신다 내가 외출했을 때 부모님은 내가 누구와 함께 있는지 대부분 알고 계신다 내가 외출했을 때 부모님은 내가 무엇을 하고 있는지 대부분 알고 계신다
물질적 요인으로 인한 스트레스	desire_stress	멋있는 옷을 입지 못해서 스트레스를 받는다 용돈이 부족해서 스트레스를 받는다 갖고 싶은 물건을 갖지 못해서 스트레스를 받는다
친구로 인한 스트레스	friend_stress	친구에게 놀림이나 무시를 당해서 스트레스를 받는다 친구들로부터 인정을 받지 못해서 스트레스를 받는다 친구에 대한 열등감 때문에 스트레스를 받는다
자기신뢰감	self_confidence	나는 내가 내린 결정을 신뢰할 수 있다 나는 내 문제를 스스로 해결할 수 있다고 믿는다 나는 내 삶을 스스로 주관하며 살고 있다
상급학교 의존도	higher_school_- dependence	자기 개발을 하기 위해서는 상급 학교 진학이 필수적이다 좋은 직업을 얻기 위해서는 상급 학교 진학이 필수적이다 좋은 조건의 배우자를 만나기 위해서는 상급 학교 진학이 필수적이다
부정적 자아존중감	neg_esteem	나는 때때로 내가 쓸모없는 사람이라고 생각한다 나는 때때로 내가 나쁜 사람이라고 생각한다 나는 대체로 내가 실패한 사람이라는 느낌을 갖는 편이다
학업으로 인한 스트레스	academic_stress	학교 성적이 좋지 않아서 스트레스를 받는다 숙제나 시험 때문에 스트레스를 받는다 공부가 지겨워서 스트레스를 받는다

Table 5: Confirmatory Factor Analysis 결과

년차	CFI	TLI	RMSEA	SRMR
1	0.927	0.916	0.041	0.042
2	0.937	0.927	0.039	0.037
3	0.930	0.919	0.039	0.043
4	0.925	0.913	0.045	0.046
5	0.934	0.924	0.042	0.045

Table 6: 종속 변수 그룹

종속 변수	원 데이터 값
안정형	대학에 진학하였고 졸업할 때까지 계속 다닐 생각이다 고등학교는 졸업하였고 현재 취업해 있는 상태이다 고등학교는 졸업하였고 현재 창업해 있는 상태이다 고등학교는 졸업하였고 현재 무보수로 가업을 돕고 있다 현재 취업 상태에서 대학생살도 병행하고 있다 현재 창업 상태에서 대학생살도 병행하고 있다 현재 무보수로 가업을 돕고 있는 상태에서 대학생살도 병행하고 있다
탐색형	대학에 진학하였고 조만간에 휴학할 생각이다 대학에 진학하였고 현재 휴학중이다 대학에 진학하였고 조만간에 중퇴할 생각이다 대학에 진학한 후 중퇴했으며 현재 다른 대학 입학 준비중이다 고등학교는 졸업하였고 현재 대학 입학 준비중이다 고등학교는 졸업하였고 현재 취업을 위해 노력중이다 대학에 진학한 후 중퇴했으며 현재 취업을 위해 노력중이다 고등학교는 졸업하였고 현재 창업을 준비하고 있다 대학에 진학한 후 중퇴했으며 현재 창업을 준비하고 있다 고등학교는 졸업하였고 현재 가사·육아를 돕고 있다 고등학교는 졸업하였고 현재 군입대 대기중이다 고등학교는 졸업하였고 현재 아무일도 하지 않고 있다

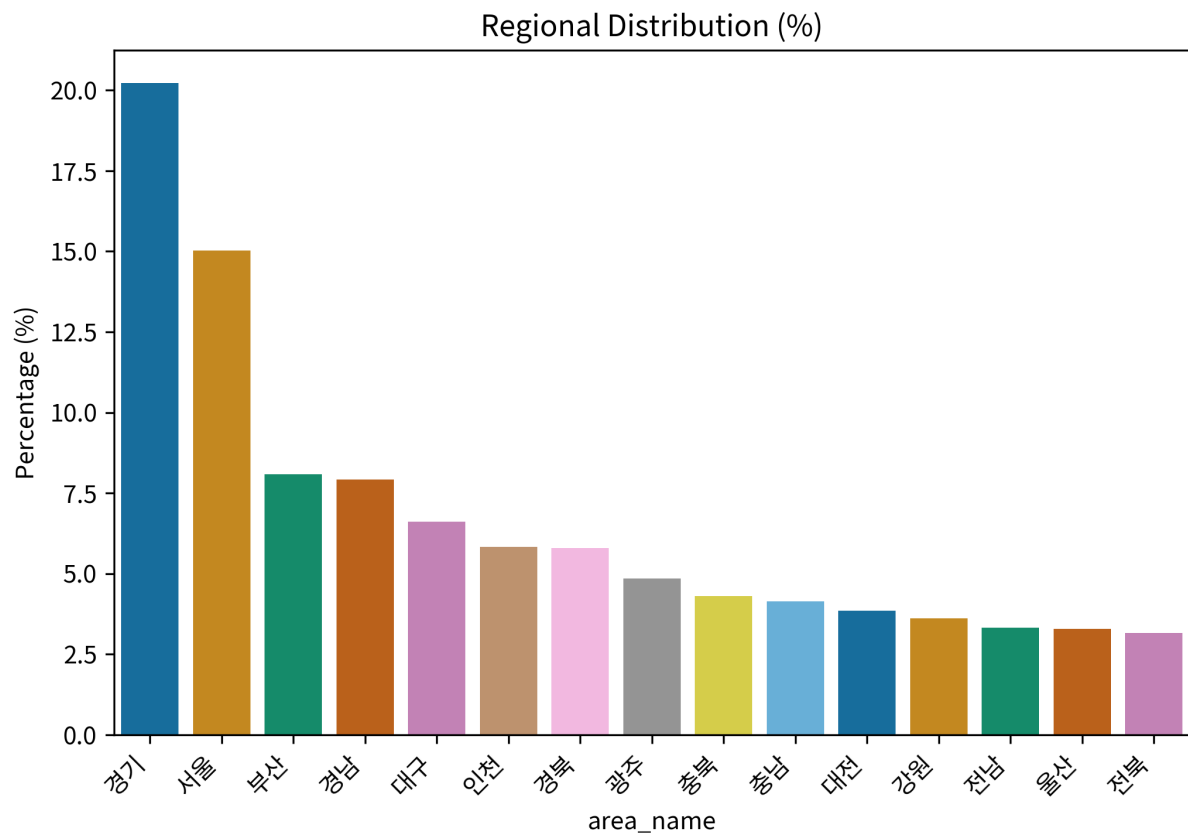
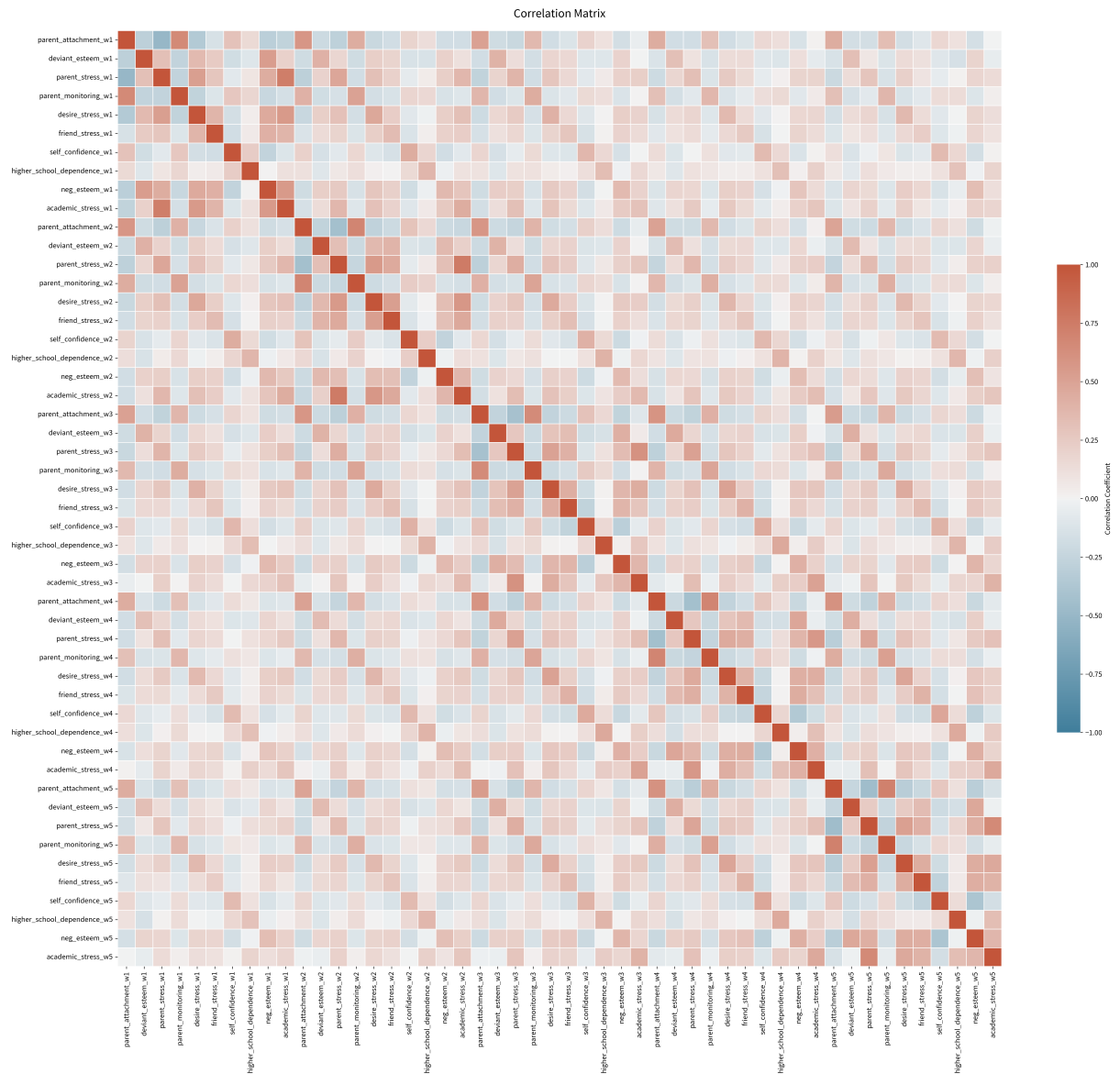


Figure 1: 지역 분포

Table 7: 가장 높은 독립변수 상관관계 10개

Variable 1	Variable 2	Correlation
parent_stress_w2	academic_stress_w2	0.757157
parent_stress_w1	academic_stress_w1	0.737240
parent_attachment_w5	parent_monitoring_w5	0.713755
parent_attachment_w4	parent_monitoring_w4	0.692874
parent_attachment_w2	parent_monitoring_w2	0.689347
parent_stress_w5	academic_stress_w5	0.663941
parent_attachment_w1	parent_monitoring_w1	0.658326
parent_attachment_w3	parent_monitoring_w3	0.655997
parent_stress_w3	academic_stress_w3	0.608676
parent_attachment_w4	parent_attachment_w5	0.604272



4 연구 과정

본 연구에서는 목표 변수를 예측하기 위해 다양한 분류 모델을 구축하고 성능을 비교 평가하였다. 데이터는 훈련 세트와 테스트 세트로 나누어졌으며, 훈련 세트에서 모델을 학습하고 테스트 세트에서 성능을 평가하였다. 훈련 세트는 70%로 설정하였고, 나머지 30%는 테스트 세트로 사용되었다.

분석에는 기준 모델로서 무작위 분류기를 시작으로, 앙상블 학습 방법인 랜덤 포레스트와 XGBoost, 그리고 전통적인 통계 모델인 로지스틱 회귀 모델이 사용되었다. 해당 모델들은 다중 공산성 문제에서 비교적 자유로워, 다양한 독립 변수들 간의 높은 상관관계가 존재하는 상황에서도 안정적인 예측 성능을 보일 것으로 기대되었다. 각 모델의 학습과정은 다음과 같다.

4.1 무작위 분류

모든 예측 모델의 성능을 비교하기 위한 가장 기본적인 기준으로 무작위 분류기를 설정하였다. 이 분류기는 훈련 데이터에서 관찰된 클래스별 비율을 기반으로 예측을 수행한다. 즉, 훈련 데이터에서 특정 클래스가 나타난 빈도에 따라 무작위로 테스트 데이터의 클래스를 할당한다. 무작위 분류기의 성능은 다른 고급 모델들의

성능이 최소한 이보다는 뛰어나야 함을 보여주는 기준선 역할을 한다.

4.2 랜덤 포레스트

본 연구에서 랜덤 포레스트 모델은 여러 개의 독립적인 결정 트리를 구축하고 각 트리의 예측을 통합하여 최종 결과를 도출하는 방식으로 학습되었다. 이 과정에서 각 트리는 원본 학습 데이터의 부트스트랩 샘플을 사용하여 학습되며, 각 분기점에서는 전체 특성의 일부분만 고려하여 최적의 분할 기준을 선택한다.

이러한 랜덤 샘플링과 랜덤 특성 선택 두 가지 무작위성이 모델의 다양성을 증가시키고 과적합을 방지하는 데 도움을 준다.

4.3 XGBoost

본 연구에서 XGBoost 모델은 순차적으로 결정 트리를 구축하는 부스팅 방식을 사용하였다. 각 트리는 이전 트리들의 오차를 최소화하는 방향으로 학습되며, 이 과정에서 손실 함수의 그래디언트를 활용한다.

이러한 순차적 학습은 이전 모델이 잘못 예측한 부분에 더욱 집중하여 전체 모델의 성능을 향상시킨다.

4.4 로지스틱 회귀

본 연구에서 로지스틱 회귀 모델은 최대 가능도 추정을 통해 학습되었다. 이 방법은 관측된 데이터가 발생할 확률을 최대화하는 모델 파라미터를 찾는 과정이다. 학습 과정에서 모델은 입력 특성의 선형 조합에 시그모이드 함수를 적용하여 각 클래스에 속할 확률을 계산한다. 본 연구에서는 안정형과 탐색형으로 구분된 두 개의 클래스가 존재한다.

4.5 모델 튜닝

모델의 최적 성능을 위해 각 알고리즘에 대해 그리드 탐색을 통한 하이퍼파라미터 튜닝을 수행하였다. 하이퍼파라미터 튜닝은 70%로 나뉘어진 훈련 세트에 대해 3-fold cross-validation을 사용하여 모델의 일반화 성능을 평가하면서 최적의 매개변수를 탐색하는 방식으로 진행되었다.

랜덤 포레스트 모델에서는 결정 트리의 개수, 트리의 최대 깊이, 분기에 사용할 특성 선택 방법, 노드 분할을 위한 최소 샘플 수, 리프 노드의 최소 샘플 수 등을 최적화하였다. 탐색 결과, 100개의 결정 트리와 특성 선택 방법으로 'log2', 트리의 최대 깊이 10, 분할을 위한 최소 샘플수 2개, 리프 노드의 최소 샘플 수 2개를 사용하는 것이 최적의 성능을 보였다.

XGBoost 모델에서는 트리의 개수, 최대 깊이, 학습률, 훈련 샘플의 서브샘플링 비율, 특성의 서브샘플링 비율 등을 최적화하였다. 최적의 매개변수 조합으로는 100개의 트리, 최대 깊이 3, 학습률 0.05, 서브샘플링 비율 0.8, 특성의 서브샘플링 비율 0.8이 선택되었다.

로지스틱 회귀 모델에서는 정규화 강도(C), 정규화 유형, solver 등을 최적화하였다. 그 결과, C 값 0.1, L2 정규화, lbfgs solver를 사용하는 모델이 가장 좋은 성능을 보였다.

이러한 하이퍼파라미터 튜닝 과정은 각 모델의 성능을 향상시키는 데 효과적이었으며, 특히 기본 설정을 사용했을 때보다 정확도가 평균 3-5% 정도 개선되는 결과를 보였다.

5 연구 결과

5.1 변수 중요도 및 회귀 계수 분석

아래의 그래프는 Random Forest의 특성 중요도, XGBoost의 특성 중요도, 그리고

로지스틱 회귀의 회귀 계수를 나타낸다. Random Forest와 XGBoost의 특성 중요도는 모델 예측에 있어 각 특성의 상대적인 기여도를 나타내며, 로지스틱 회귀의 회귀 계수는 각 특성이 목표 변수에 미치는 영향의 방향성과 크기를 정량적으로 제시한다. 이러한 결과들을 종합적으로 살펴봄으로써, 어떤 특성이 청소년의 진로 유형 선택에 주요한 영향을 미치는지 다각적으로 파악할 수 있다.

로지스틱 회귀 분석 결과에서 회귀 계수는 각 특성이 목표 변수인 진로 유형에 미치는 영향의 방향성을 나타낸다. 양(+)의 계수는 해당 특성이 탐색형 진로 선택 가능성을 높이는 요인임을, 음(-)의 계수는 안정형 진로 선택 가능성을 높이는 요인임을 의미한다. 본 해석에서는 유의수준 95% ($p \leq 0.05$)를 기준으로 통계적 유의성을 보인 변수들을 중심으로 논의한다.

먼저, 탐색형 진로 패턴을 유의하게 예측하는 요인으로는 '중학교 2학년의 물질적 스트레스(desire_stress_w1)', '중학교 3학년의 물질적 스트레스(desire_stress_w2)', 그리고 '고등학교 2학년의 부정적 자아존중감(neg_esteem_w4)'이 확인되었다. 이 변수들은 Random Forest 모델에서 각각 2위(89.19), 19위(67.66), 8위(79.71)를, XGBoost 모델에서는 각각 14위(74.24), 16위(73.58), 2위(89.68)의 중요도를 보여, 트리 모델에서도 그 영향력이 작지 않음을 알 수 있다. 중학교 시기에 경험하는 경제적 스트레스는 청소년으로 하여금 조기 경제적 자립에 대한 압박감이나 강한 동기를 느끼게 하여, 장기적 안정성이 보장되지 않더라도 다양한 시도를 하거나 기존 경로에서 벗어나는 탐색적 행동을 촉진할 수 있음을 시사한다. 또한, 고등학교 2학년의 높은 부정적 자아존중감은 기존 학업 경로에 대한 확신 부족이나 부적응감으로 이어져, 다양한 가능성을 시도하며 자신에게 맞는 길을 찾으려는 탐색 과정이나 진로 결정 유예 경향으로 이어질 수 있음을 보여준다.

반대로, 안정형 진로 패턴을 유의하게 예측하는 요인으로는 '고등학교 1학년의

자기신뢰감(self_confidence_w3)', '고등학교 2학년의 자기신뢰감(self_confidence_w4)', '고등학교 2학년의 친구로 인한 스트레스(friend_stress_w4)', 그리고 '고등학교 3학년의 상급학교 의존도(higher_school_dependence_w5)'가 확인되었다. 이 변수들은 Random Forest 모델에서 각각 21위(66.83), 1위(100.00), 5위(82.51), 22위(66.33)를, XGBoost 모델에서는 각각 27위(67.71), 3위(87.71), 6위(83.60), 19위(72.31)의 중요도를 나타냈다. 이는 고등학교 시기의 높은 자기신뢰감이 자신의 능력에 대한 믿음을 바탕으로 설정한 목표를 꾸준히 추구하게 하여 안정적인 진로 선택으로 이어질 가능성을 시사한다. 고등학교 2학년 시기의 친구 관계에서 오는 스트레스는, 대인 관계의 어려움으로 인해 오히려 자신의 통제하에 있는 학업이나 개인적 목표 달성에 더 집중하게 함으로써 안정형 진로 선택 가능성을 높이는 것으로 해석할 수 있다. 고등학교 3학년의 높은 상급학교 의존도는 상급학교 진학을 자기계발과 성공적인 미래(좋은 직업, 좋은 배우자 등)를 위한 필수적인 단계로 여기는 인식을 나타내는 문항과 연관되어 있다. 이는 곧 뚜렷한 장기적 목표를 설정하고, 이를 달성하기 위해 계획된 경로를 일관되게 추구하며 끝까지 나아가려는 의지와 맞닿아 있어, 안정형 진로 선택의 중요한 예측 요인이 됨을 시사한다.

무작위 분류기(Random Classifier)보다는 향상된 성능을 보였으나, 전반적인 예측 정확도와 특히 '탐색형' 범주에 대한 분류 성능(예: 재현율, F1-점수)이 기대에 미치지 못하여 제한적인 예측력을 확인하였다.

5.2 모형 평가 결과

모형 성능 평가는 정확도(accuracy), 정밀도(precision), 재현율(recall), F1-score 등 다양한 지표를 사용하여 수행되었다. 각 모형의 성능은 아래 [그래프](#)와 [ROC curve](#)와 같다.

분석 결과, 본 연구에서 활용된 청소년기 심리·정서적 특성 변수들은 만 18-19세 시점의 진로 상태(안정형 vs. 탐색형)를 명확하게 구분하는 데 있어 높은 수준의 예측력을 나타내지는 못했다. 세 가지 예측 모형(Random Forest, XGBoost, Logistic Regression) 모두

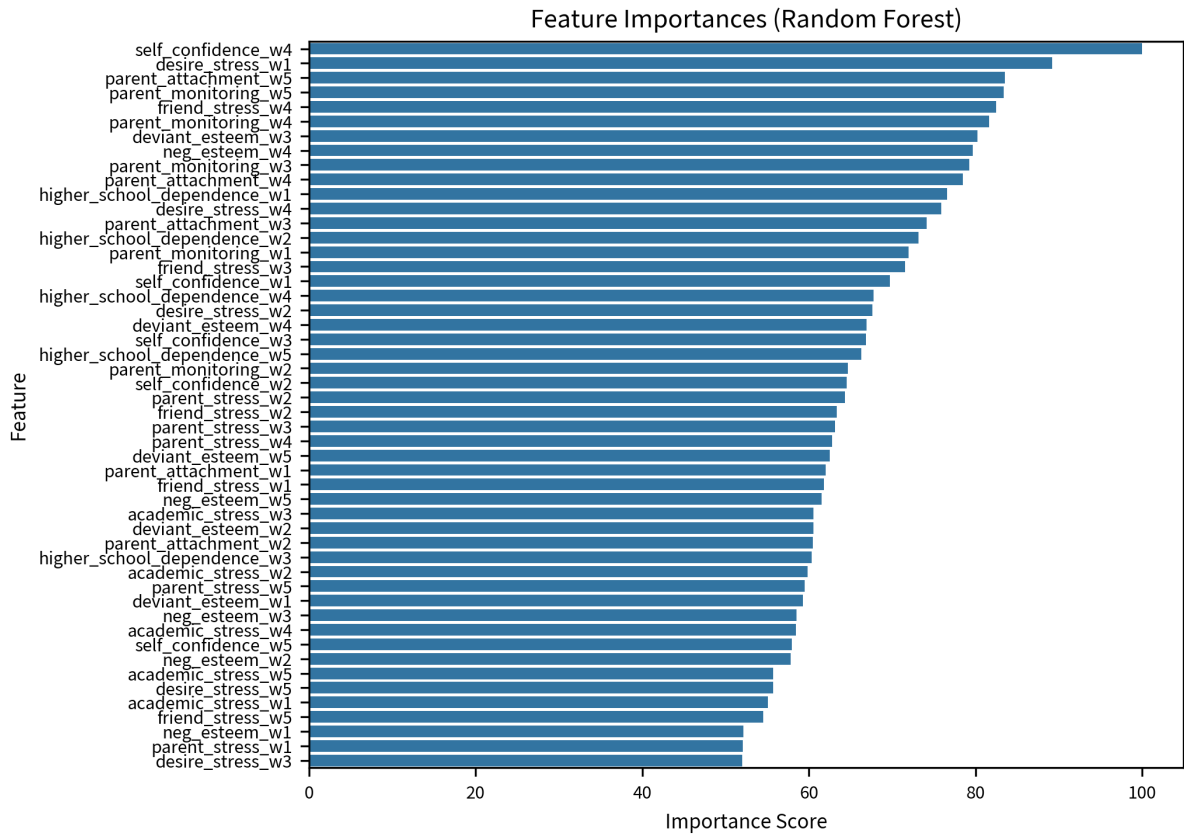


Figure 3: 랜덤 포레스트 변수 중요도

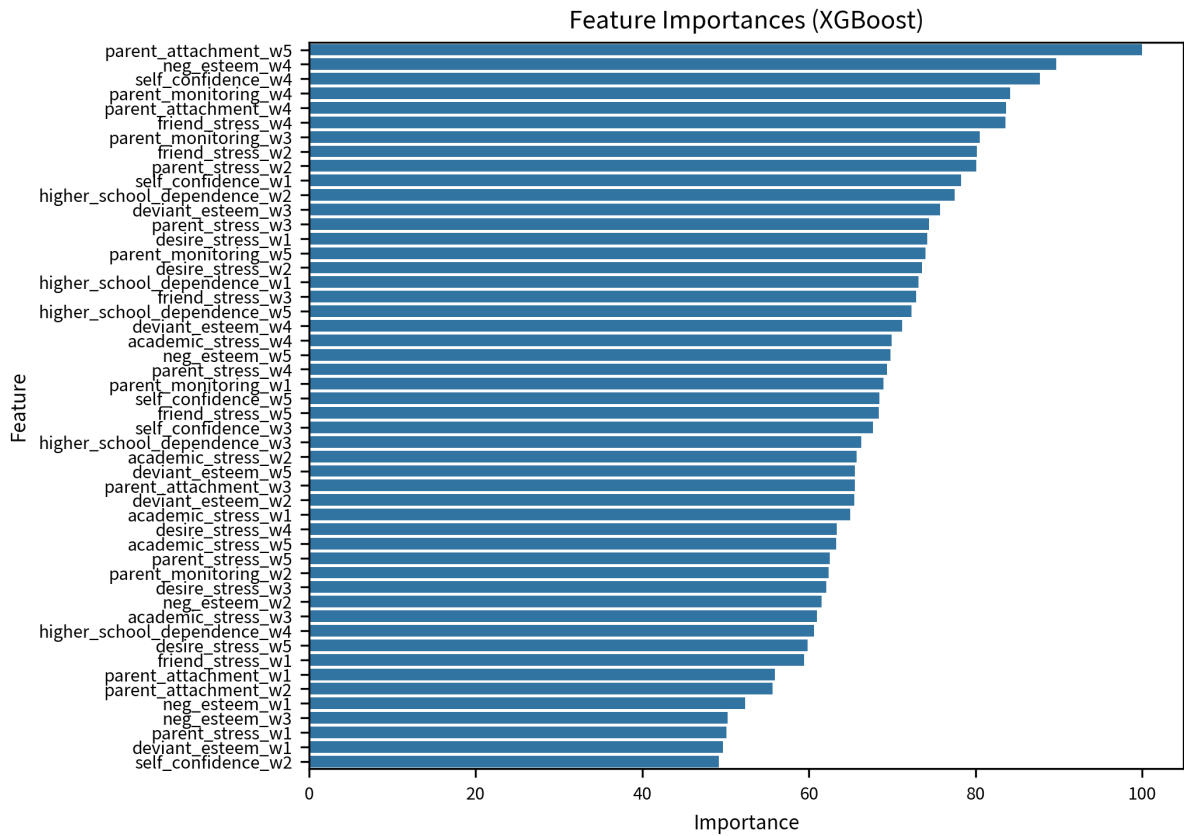


Figure 4: XGBoost 변수 중요도

Table 8: Logistic Regression 계수 (p-value < 0.05)

변수명	계수	p value
desire_stress_w1	0.499608	0.000015
self_confidence_w4	-0.332916	0.010294
neg_esteem_w4	0.263964	0.031803
desire_stress_w2	0.255400	0.018759
self_confidence_w3	-0.254064	0.030035
friend_stress_w4	-0.247249	0.024386
higher_school_dependence_w5	-0.164889	0.024659

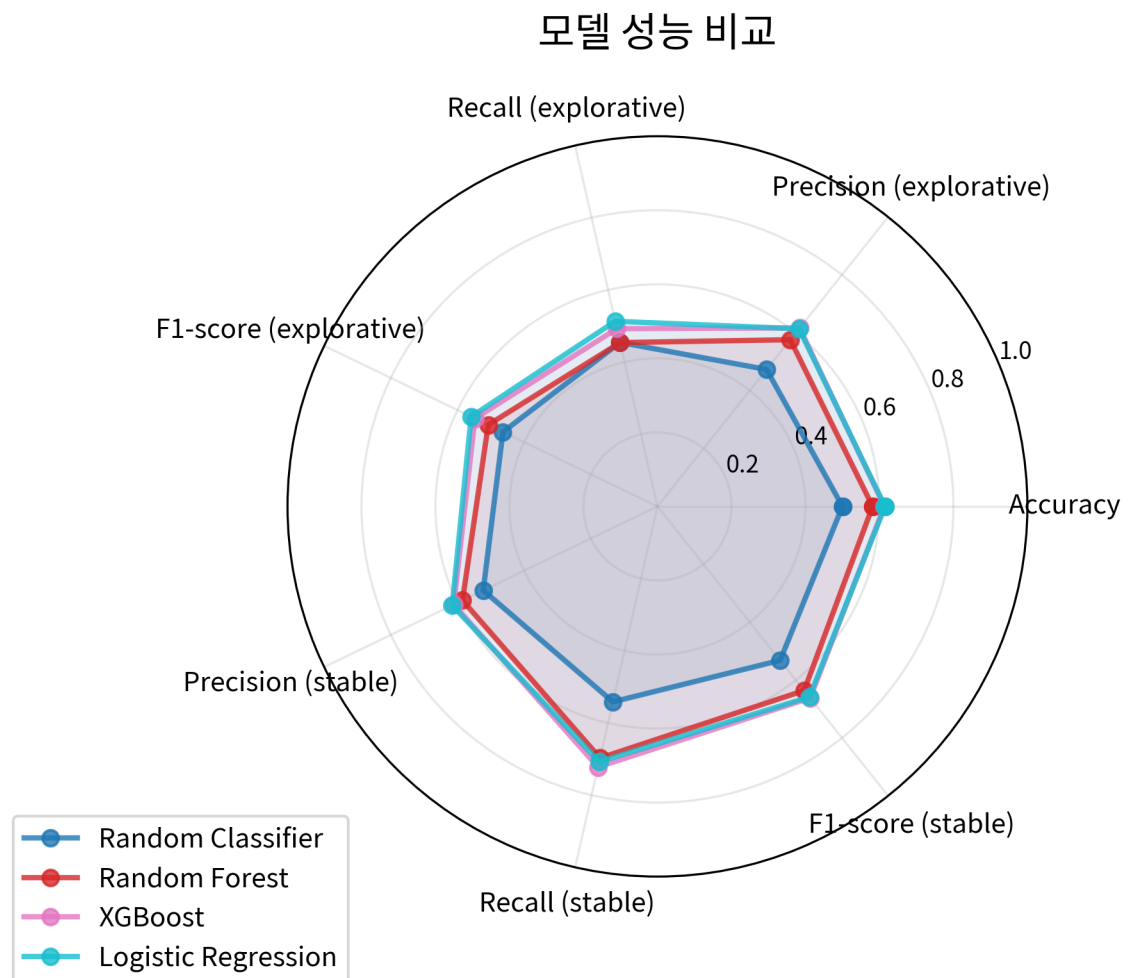


Figure 5: 모델 성능 비교

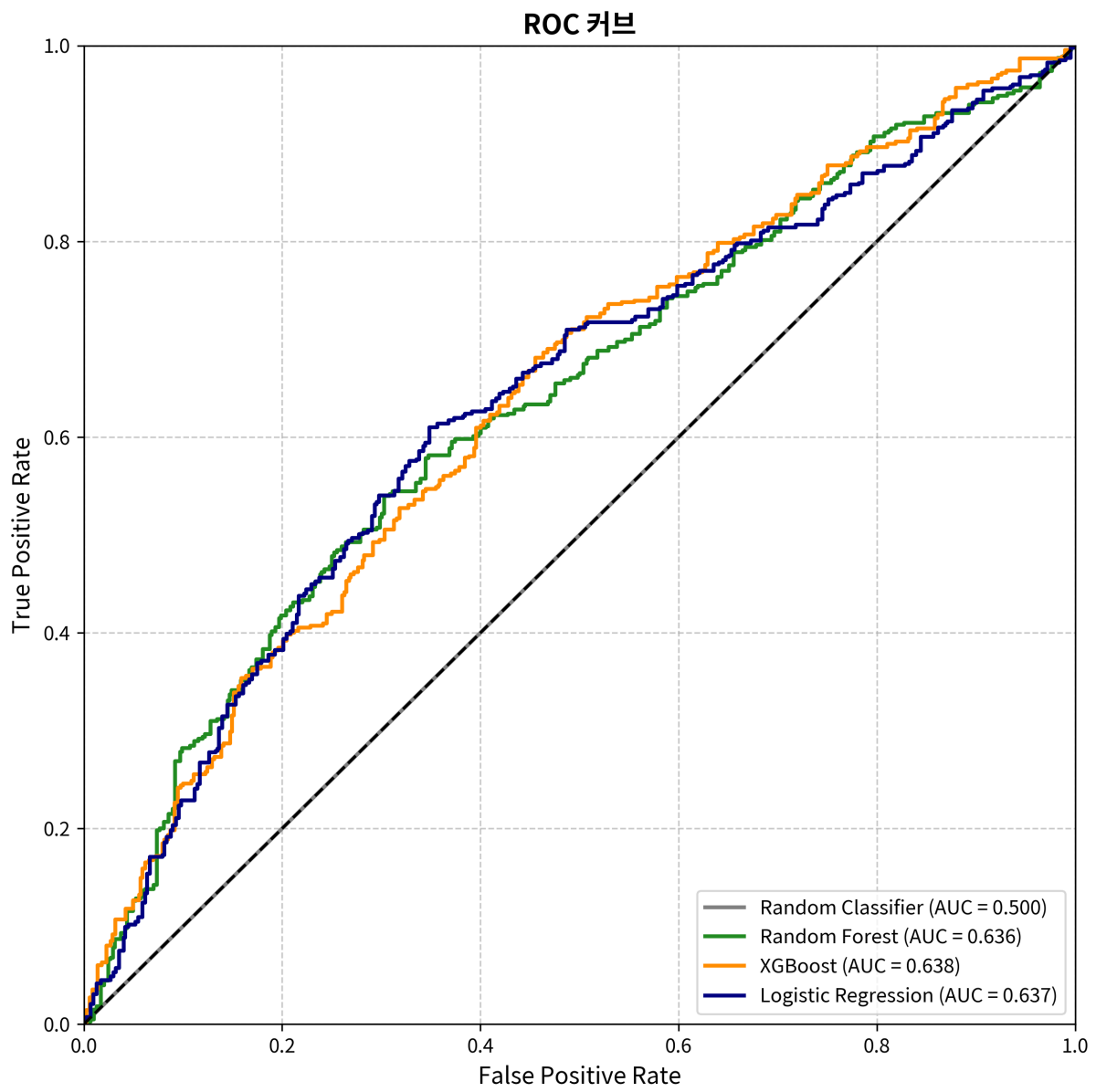


Figure 6: ROC 그래프

6 결론

6.1 연구 의의와 발전 방향

본 연구는 청소년의 진로 유형을 예측하고자 다양한 개인 내적 및 환경적 요인을 다각도로 고려하였으며, 특히 중학교 2학년년부터 고등학교 3학년까지 6년간의 패널 데이터를 활용하여 시간에 따른 이들 요인의 영향력 변화 가능성을 탐색했다는 점에서 초기적인 의의를 찾을 수 있다. 비록 개발된 예측 모델의 성능이 기대만큼 높지 않았고 분석 과정에서 여러 한계점이 드러났으나, 이러한 시도 자체가 청소년 진로 발달 과정의 복잡성을 재확인하고 향후 관련 연구를 위한 중요한 시사점을 제공한다는 점에서 의미가 있다고 판단한다.

본 연구에서 확인된 주요 한계점과 이를 보완하여 향후 연구를 발전시킬 수 있는 방향은 다음과 같다.

첫째, 모델의 예측 성능과 관련된 한계점이다. 본 연구에서 개발된 모델들은 무작위 분류에 비해 향상된 결과를 보였음에도 불구하고, 실제 청소년의 다양한 진로 유형을 명확하게 구분하는 데에는 미흡한 점이 있었다. 이러한 한계는 본 연구에서 활용된 표본의 크기가 상대적으로 작았다는 점, 표본 수집 과정에서 특정 지역으로 편중되었을 가능성으로 인한 대표성 확보의 어려움, 또는 본 연구에서 주로 탐색된 분석 모델들이 가진 내재적 한계나 다른 분석 전략 탐색의 여지 등에서 기인했을 수 있다. 따라서 향후 연구에서는 더 큰 규모와 함께 지역 및 인구학적 특성을 고려한 균형 잡힌 대표성 있는 표본을 확보하는 것이 중요하다. 더불어, 본 연구에서 관찰된 바와 같이 특정 진로 유형에 대한 예측 성능이 상대적으로 낮게 나타난 문제를 해결하기 위해, 소수 클래스의 데이터를 증강하는 오버샘플링 기법의 적용을 신중히 고려해볼 수 있다. 또한, 본 연구에서 활용된 로지스틱 회귀, 랜덤 포레스트, XGBoost 모델 외에도 다양한 머신러닝 알고리즘(예: 서포트 벡터 머신, 딥러닝 모델 등)의 적용 가능성을

탐색하거나, 통계적 분석을 보완할 수 있는 질적 연구(예: 심층 인터뷰, 사례 연구)를 병행하여 진로 선택의 맥락적 이해를 심화시키고 모델의 설명력과 예측력을 제고할 필요가 있다.

둘째, 종속변수 정의 및 분류의 단순화 문제이다. 본 연구에서는 청소년의 진로 상태를 '안정형'과 '탐색형'이라는 이분법적 범주로 단순화하여 분석을 진행하였다. 그러나 실제 청소년의 진로 발달 상태와 준비 과정은 훨씬 더 다양하고 복잡한 스펙트럼을 가질 수 있으며, 이러한 이분화는 현실의 다층적인 복잡성을 충분히 반영하지 못했을 가능성이 있다. 따라서 향후 연구에서는 현재의 이진 분류를 넘어 보다 세분화된 다중 클래스로 종속변수를 확장하거나, 진로 준비도나 진로 성숙도와 같은 연속형 변수를 예측하는 회귀 문제로 전환하여 분석의 폭과 깊이를 더하는 방안을 고려할 수 있다.

셋째, 독립변수 선택 및 활용에 대한 성찰이다. 본 연구에서는 광범위한 변수를 탐색적으로 모델에 포함시키다 보니, 일부 변수는 이론적 중요도에 비해 실제 예측 기여도가 낮게 나타나거나 다른 변수와의 개념적 중복 가능성이 확인되었다. 또한, 다수의 독립변수로 인해 모델 해석의 복잡성이 증가하고, 각 변수의 고유한 영향력을 명료하게 파악하는 데 어려움을 겪었다. 향후 연구에서는 이론적 논의와 선행연구 고찰을 바탕으로 진로 유형 예측에 핵심적인 역할을 할 것으로 기대되는 변수를 보다 정교하게 선별하고, 모델의 간명성을 높이는 방향으로 개선이 이루어져야 할 것이다. 이는 변수 간 관계 및 영향력에 대한 더 명확한 이해를 도울 뿐만 아니라, 모델의 일반화 가능성 또한 높일 수 있을 것이다.

결론적으로, 본 연구는 여러 한계점에도 불구하고 청소년 진로 발달에 대한 종단적 접근의 필요성을 강조하고, 다양한 심리사회적 변인들의 복합적인 영향을 탐색적으로 확인했다는 점에서 그 의미를 찾을 수 있다. 위에 제시된 발전 방향을 통해 향후 연구에서는 보다 정교한 모델 구축과 심층적인 분석이 이루어지기를 기대하며, 이를 통해 청소년의 진로 발달을 보다 깊이 있게 이해하고

실질적인 지원 방안을 마련하는 데 기여할 수 있기를 바란다.

7 참고문헌

- [한국 아동 청소년 청년 데이터 아카이브](#)
- [김미소. \(2025\). 대학생의 자아존중감과 자기효능감이 진로 성숙도에 미치는 영향 \(석사학위논문\). 제주대학교 대학원](#)
- [강선아. \(2023\). 자아존중감이 진로성숙도에 미치는 영향에 관한 연구: 다변량 잠재성장모형의 종단연구 중심으로 \(박사학위논문\). 송실대학교 대학원](#)
- [이윤희. \(2013\). 청소년 진로·직업역량에 영향을 미치는 요인에 관한 연구 \(석사학위논문\). 경기대학교 교육대학원, 청소년지도상담교육전공](#)
- [장슬비. \(2023\). 청소년의 진로미결정이 우울감에 미치는 영향: 그릿\(GRIT\)의 조절효과 \(석사학위논문\). 연세대학교 사회복지대학원, 사회복지학 전공](#)