

## 랜덤포레스트를 활용한 특성화고 학생들의 졸업 후 진로결정 예측요인 탐색

김 하 정<sup>1)</sup> · 원 효 현<sup>2)</sup>

### 요 약

본 연구는 한국교육고용패널 II(Korean Education & Employment Panel: KEEP II)의 1차, 2차, 3차년도 자료를 활용하여 특성화고등학교 학생들의 졸업 후 진로 결정에 미치는 주요한 영향요인이 무엇인지를 탐색하고, 학생들의 진로결정에 도움을 줄 수 있는 방안 마련 및 현장에서의 교육적 활용성을 높이고자 하였다.

연구 결과, 특성화고 학생들의 졸업 후 진로결정에 영향을 주는 요인으로 '고등학교 3학년 시기의 희망교육수준'이 가장 높은 예측변인으로 나타났으며, 다음으로 '특성화고 선택 이유', '고등학교 2학년 시기의 희망교육수준', '3학년 1학기 수학 내신 등급', '3학년 때 경험한 취업상담'의 순으로 중요도가 높은 것으로 나타났다. 또한 중요도 지수 상위 30개 변수에 학업 및 성적 관련 변수가 총 10개로 학생들의 졸업 후 진로 결정을 예측하는 중요한 변수로 고려되었으며, 성별, 키, 지역 규모, 여가시간, 행복도 등 개인특성 관련 변수(7개)와 교육포부와 삶의 관점 등 진로에 대한 인식 관련 변수(5개) 및 가정 특성 변수(4)와 학교특성 변수(4)도 주요한 예측 변수로 도출되었다. 이러한 결과를 바탕으로 특성화고 졸업생들의 진로결정에 도움을 줄 수 있는 교육적 방안 마련을 위한 시사점을 논의하였다.

## I. 서론

한편, 국가 통계(KOSIS)에 따르면 2022년의 평균 청년(15세~29세) 실업률은 5.2%로 전체 평균 실업률 3.0%을 크게 웃돌고 있으며, 상반기 기준 청년 체감실업률은 19.9%로 다른 연령대보다 월등히 높았다. 또한 2016~2026 중장기 인력수급전망(고용노동부, 2017)에 따르면, 대졸자 이상 학력자의 초과공급은 75만 명, 고졸 이하 학력자에 대한 초과수요는 약 113만 명이 발생할 것으로 예상된다. 즉 대졸자 이상의 고학력자는 구직난을 겪고 있는 동시에 중소기업 현장에서는 인력난을 겪고 있는 노동시장의 구조적 미스매치가 발생하면서 노동시장의 수급 불균형이 악화되는 상황이다(강민수, 2021). 이처럼 높은 청년 실업률은 인력수급 불일치와 학력 차별 등의 문제와 결부되어 있다(홍성표, 정진철, 2016).

이에 정부는 고학년 청년실업률의 부담을 경감시키고, 고교 단계의 취업을 위한 직업교육을 강조하고자 중등 직업교육 정책을 취업 중심 체제로 전환하고, 특성화고 학생들의 취업 관련

1) 부경대학교 연구교수

2) 부경대학교 교수

역량 강화, 노동시장의 요구에 맞는 인력 양성 등 교육과 노동시장 간의 원활한 이행구조 확립을 위해 다양한 정책을 추진해왔다(전현정, 정혜원, 2018). 예컨대 중등직업교육의 개혁 방안으로 한국형 마이스터고 육성계획, 고교 직업교육 선진화, 고졸 취업 활성화, 산학일체형 직업교육 활성화, 선취업-후진학 정책 등이 추진되었으며, 직업계고등학교의 취업률은 2010년 19.2%에서 2020년 50.7%로 비약적으로 상승하였다(교육부, 2019; 이정민 외, 2019; 정송, 2020). 이러한 노력은 특성화고 체질 개선을 통해 직업교육을 혁신하고 직업계 고등학교에 대한 학생, 학부모, 기업들의 부정적인 인식 개선을 꾀하려는 목적 또한 담고 있다.

그러나 중학교 성적이 좋지 않은 학생들이 특성화고등학교에 진학한다는 사회적 편견과 함께 고졸자에 대한 학력 차별은 아직 우리 사회에 존재하고 있다(김정숙, 나민주, 2015; 이휘인, 김현철, 2015). 이로 인해 능력위주의 채용 사례는 소수에 불과하다는 인식이 지배적이며, 고졸 취업자들을 위한 양질의 일자리 부족과 열악한 근무환경은 초기 직장 적응과정에서의 이탈 또는 변경의 원인으로 언급되고 있다(노경란, 변정현, 2016; 두민영, 우버들, 박윤희, 허영준, 오계택, 2013; 신창호, 2018; 이병욱 외, 2013). 이러한 직업교육에 대한 사회적 편견, 기술직에 대한 좋은 일자리 부족, 경력 개발 기회 부족 등으로 인해 특성화고등학교 학생들은 재학 중 여러 차례 진로결정을 수정하기도 하며, 대학진학을 선택하는 비율이 증가하는 원인이 되기도 하였다(김홍기, 김지훈, 김양현, 2017). 또한 마이스터고 학생들과 달리 특성화고 학생들의 경우 취업과 대학 진학의 선택지를 모두 가지고 있기 때문에 진로결정에 있어 비교적 복잡한 양상을 보이며, 일반고에 비해 심각한 갈등을 겪는 것으로 알려져 있다(강민수, 2021, 재인용). 특히 우리나라는 대학 졸업자 중심으로 취업시장이 형성되었고, 특성화고 동일계 학생에 대한 대학 진학 기회 증가로 인해 특성화 고등학교가 ‘특정분야 인재 및 전문 직업인 양성을 위한 취업 중심·산학협력 학교’라는 본래의 설립 목적에 맞지 않게 운영되고 있다는 지적이 있다(이휘인, 김현철, 2015). 실제 한국교육개발원(2022)의 교육통계자료에 의하면 2018년 대비 2021년의 특성화고 졸업생 취업자 수는 3만 8천여 명에서 1만 8천여 명으로 절반 이상 감소하였으며, 2017년 32.8%의 대학 진학률은 2020년 기준 47.4%까지 증가했다.

특성화고 학생들의 진로결정에 관한 초기 연구들을 살펴보면 성별, 학업성취, 교육포부, 가정의 사회경제적 배경이나 가정생활만족도 등 개인특성 변인들과(김기현, 방하남, 2004; 이지혜, 정철영, 2010; 채창균, 2009) 학교생활적응, 시설, 진로지도 등 학교 맥락변인(김경근, 변수용, 2006; 최동선, 2008) 등이 특성화고 학생들의 진로결정과 관련이 있다고 보고되었다. 최근에는 주로 특성화고와 다른 고교 유형 간 진로결정 수준 비교 연구(길혜지 외, 2018; 김수정 외, 2016, 최보금, 2020), 진로성숙도 및 진로준비행동의 영향 요인 탐색(김진원, 모영민, 2019, 윤초희, 안태희, 2015, 최보금, 2020), 진로결정에 영향을 주는 변인 탐색 연구(김수정 외, 2013; 김정란, 2006; 오석영, 2012; 이명훈, 2016; 이은경, 2014; 홍성표, 정진철, 2016) 등을 중심으로 취업 또는 진학에 영향을 미치는 사회적·환경적 요인들과 노동시장 정착 가능성을 설명하는 요인들을 다루고 있다. 이들 연구들은 특성화고 학생들의 진로결정에 영향을 미치는 다양한 요인들을 탐색했다는 점에서 의의가 있다. 그러나 개별 요인들 간의 인과관계를 밝힌 연구가

대부분이며, 다양한 변인들을 투입하여 진로결정에 영향을 미치는 우선 요인이 무엇인지에 대해 밝힌 최근의 연구는 계열 및 지역에 한정된 연구라는 제한점이 있다.

이에 본 연구에서는 종속변수와 예측변수 간의 비선형성 및 예측변수 간의 상호작용을 고려한 추정 결과를 도출할 수 있으며, 전통적 통계방법의 한계를 보완하고 동시에 다양한 변수를 고려할 수 있다는 점에서 랜덤포레스트(random forest) 분석방법을 활용함으로써 그 한계를 극복하고자 한다. 이를 통해 선행연구에서 밝힌 개인, 가정, 학교 차원의 다양한 변수들을 함께 투입하여 여러 변수들 가운데 특성화고등학교 학생들의 졸업 후 진로 결정에 미치는 주요한 영향요인이 무엇인지를 탐색하여 현장에서의 교육적 활용성을 높이고자 한다.

## II. 이론적 배경

### 1. 진로결정의 개념

진로결정의 개념은 국내외의 여러 학자들에 의해 다양하게 논의되고 있다. 개인의 특성과 직업의 요구가 적절히 충족될 때 진로가 결정된다는 이성적 판단 관점의 주장(Gati, 1986; Parsons, 1990)과, 개인의 학습경험과 환경적 조건, 사건 등 다양하고 복잡한 요소와 관련한 의식적 노력이나 시행착오를 통해 진로가 결정된다는 계획된 우연의 관점(Mitchell et al., 1999)의 주장이 있으며, Krumboltz & Drumboltz(1976)은 진로결정을 진로에 대한 사회적·직접적 강화 또는 모델화의 결과로 선호적 진로를 선택하는 능력으로 규명하였다. 또한 학교교육에 있어 진로결정은 교육에 기반하고 직업을 준비하는 과정을 포함한 개념으로, 학교 교육과정을 통해 자신에게 적합한 직업을 찾는 능력을 의미하는 것이 일반적이며(Crites, 1981), 최은영(2011)은 진로결정을 자신의 적성, 능력, 흥미에 대한 분명한 이해와 함께 다양한 직업에서 요구하는 기술들을 상호 비교함으로써 자신에게 가장 적합하고 실행 가능한 대안을 선택하는 과정이라고 정의하였다. 즉, 학교 교육을 기반으로 하고 있는 현대 사회에서 진로 결정은 교육을 바탕으로 다양한 사회적·환경적 영향을 받게 된다(전현정, 정혜원, 2019). 이러한 진로 결정은 개인마다 차이가 있으며, 진로결정 과정은 개인의 인지적·환경적 요인들과 밀접한 관련이 있다(이휘인, 김현철, 2015 재인용). 즉, 고등학생의 직업가치관은 중학교시기에 형성된 가치관의 영향이며(김경년, 2011), 일반계 고등학생의 진로결정은 대학 진학 시에는 전공 선택의 형태로, 노동시장 진입 시에는 직업선택을 통해 구체화되고(황여정, 2007), 특성화고 학생들의 경우 취업 또는 진학을 선택하는 과정이 진로 결정의 과정으로 여러 변인들의 영향을 받아서 결정된다(김경년, 2011).

### 2. 특성화고 학생들의 진로결정 요인에 관한 선행연구 탐색

특성화고 학생들의 진로결정 요인을 분석한 연구는 <표 1>과 같다.

〈표 1〉 선행연구에 제시된 진로결정 관련 변인

연구자	내용
신희경, 김우영 (2005)	일반고: 성적, 내적동기, 자아관, 부모와의 대화시간, 부모와 진로결정 정보 찾기 특성화고: 내적동기, 교직경력, 교사 성별, 연령, 진로 관련 강연 경험
오현석 외(2010)	개인변인: 성별, 학업성취, 적성인지 가정변인: 가정만족도, 결손가정여부, 가족활동 학교변인: 계열, 학교시설, 학과만족, 진로교과, 진로체험
오석영(2012)	개인변인: 성적, 자기효능감, 미래직업결정, 적성 가정변인: 대화 학교변인: 학교생활만족도, 진로와 직업수업, 학과안내, 진로검사, 진로상담, 진로관심교사
배정수, 송병국 (2015)	특성화고: 자기이해, 진로 및 직업정보 부족, 긍정양육태도, 친구지지, 대안선택 효능감 일반고: 자기이해부족, 진로 및 직업정보 부족, 미래계획효능감, 부모 교육수준, 성역할갈등 및 자기 확신 부족, 특강직업인 만남 특목고: 자기이해부족, 미래계획효능감, 부모 교육수준, 각종검사경험, 친구지지, 성역할갈등 및 자기 확신부족, 부모지지
이명훈(2015)	직업정보 부족, 자기 명확성 부족, 필요성 인식 부족 교과지도 만족, 진로교육 만족, 중학교 때 현재 학교 희망 여부, 성별, 학업성적, 계열, 부모의 진로관련 모델링, 부모의 정서적 지지
이휘인, 김현철 (2015)	부모의 기대, 진로활동경험, 성적
김지효(2018)	2013년: 성별, 사교육 참여정도, 학업성취, 동아리 활동, 부모의 진학정보 인식수준, 학교장의 진로교육정책에 대한 필요성 인식수준, 취업정보 제공 및 습득, 진로체험 운영 2016년: 성별, 사교육 참여정도, 부모의 진학정보 인식수준, 진로상담, 체험기반 진로프로그램 참여 경험

일반고와 실업계고 학생의 진로결정에 영향을 미치는 요인을 개인, 가정, 학교변인으로 구분하여 분석한 신희경, 김우영(2005)의 연구에서는 일반고 학생의 경우 학생의 성적과 내적 동기, 자아관과 같은 개인변인과 부모와의 대화시간 및 부모와 진로결정 정보 찾기 등 가정변인이 주요한 변인으로 나타났다. 특성화고 학생의 경우 개인변인인 내적동기 외에 교직경력, 교사 성별, 연령, 진로 강연 경험 등의 학교변인이 진로결정에 영향을 주는 주요한 변인이라고 보고하였다. 오현석 외(2005)의 연구도 일반고와 특성화고 학생들의 진로결정 수준을 비교하여 분석하였는데, 성별, 학업성취, 적성인지 등의 개인요인과 가족활동, 진로교과, 진로체험 등이 유의미한 영향을 나타냈으며, 학업성적이 높은 특성화고 학생들의 대학진학 경향이 높게 나타났다. 오석영(2012)의 연구 결과 학과 및 학교안내 참여 학생들이 대학 진학보다 취업을 결정하는 경향을 보였으며, 그 외 진로상담의 취업 결정 효과를 밝혔다. 배정수와 송병국(2015)은 의사결정나무 분석기법을 활용하여 특성화고, 일반고, 특목고 집단의 진로결정 요인 우선순위를 탐색하였으며, 특성화고 학생들의 경우 자기이해부족, 진로 및 직업정보 부족, 긍정양육태도, 친구지지, 대안선택효능감이 진로결정에 정적·부적 영향을 주는 것으로 나타났다. 이와 유사

하계 이명훈(2015)의 연구에서도 직업정보, 자기 명확성, 필요성 인식, 교과지도 만족, 진로교육 만족, 부모의 지지 등이 관련요인으로 검증되었으며, 교과 및 진로지도 만족과 현재 학교 희망 여부 등이 진로결정과 상관이 있었다. 특성화고 학생들의 진로결정(취업, 진학, 미결정) 영향 요인을 탐색한 이휘인, 김현철(2015)은 부모의 기대, 진로활동경험, 성적이 진로결정에 영향을 준다고 보고하였다. 한편, 김지효(2018)은 진로교육 정책의 변화과정에 따라 특성화고 졸업생의 취업결정 및 성과에 미치는 영향요인을 분석하고자 하였으며, 그 결과 2013년에는 성별, 사교육 참여정도, 학업성취, 동아리 활동, 부모의 진학정보 인식수준, 학교장의 진로교육정책에 대한 필요성 인식수준, 취업정보 제공 및 습득, 진로체험 운영 등이, 2016년에는 성별, 사교육 참여정도, 부모의 진학정보 인식수준, 진로상담, 체험기반 진로프로그램 참여 경험 등이 취업결정에 유의한 영향을 주었다고 보고하였다.

진학, 취업, 미결정 집단별로 진로결정 요인을 분석한 선행연구를 살펴보면, 가정의 사회경제적 배경, 부모와의 유대감, 부모의 교육포부, 학업성취, 미래직업에 대한 포부, 교사의 열의, 학교 정책, 교사와의 관계 등이 대학 진학 결정에 영향을 준다는 결과가 보고되었다(오석영, 2012 재인용). 그러나 학교교육보다 가정 및 개인변인이 대학 진학에 영향을 미친다(최동선, 2007; Coleman, 1990)라는 결과도 발견되었다. 졸업 후 취업을 결정하는 요인으로는 경제적 이유, 사회경험 욕구, 학업에 대한 열의 부족이 영향을 준다는 결과(조용 외, 2011; 조근제, 김선진, 2009)가 보고되었으며, 취업관련 경험은 취업을 결정하는 정적 요인으로(김성환, 전용석, 2004; 채창균, 2009), 교육포부, 학교생활적응, 자기효능감은 부적 요인으로(안선영, 이경상, 2009) 나타났다. 진로 미결정 요인으로는 성별, 성적, 자아개념, 중요한 타인들의 영향, 미래에 대한 확신(Gordon, 1981), 직업에 대한 정보와 자신감 및 진로에 대한 관심 부족(정윤경, 2000), 성적, 구직효능감, 취업경험, 가구 수입, 부의 교육수준 등의 부적 효과(이지혜, 정철영, 2010) 등이 보고되고 있다.

### Ⅲ. 연구방법

#### 1. 분석 대상

본 연구는 한국직업능력연구원의 한국교육고용패널(KEEP II) 자료 중 1차년도(2016), 2차년도(2017), 3차년도(2019)자료를 활용하였다. 분석대상은 1차년도 당시 고등학교 2학년, 2차년도 당시 고등학교 3학년, 그리고 졸업 1년 후에 해당하는 3차년도 조사 참여자들 중 특성화고 학생 자료이며, 1차년도 조사에 응답하지 않아 고교유형 확인이 어려운 학생들과, 1, 2차년도 응답자 중 3차년도 조사에 참여하지 않아 졸업 후의 진로 확인이 불가능한 학생들을 분석대상에서 제외한 1,775명을 최종 분석 대상으로 선정하였다.

## 2. 변수 설명

### 가. 결과변수

본 연구의 결과변수는 KEEP II 1차년도(2016), 2차년도(2017), 3차년도(2019) 데이터를 활용하여 다음과 같은 기준으로 ‘취업’, ‘진학’, ‘기타’로 분류하였다. 우선 1차년도 응답 학생 중 특성화고 재학생을 대상으로 3차년도 현재학력에 대한 응답이 대학 재학 중인 경우 ‘진학’으로 분류하였으며, 고등학교가 최종학력인 학생 데이터 중 현재 일자리가 있는 학생은 ‘취업’으로, 그렇지 않을 경우 ‘기타’로 분류하였다. 고교유형과 졸업 후 진로가 확인되지 않는 학생 데이터는 제외하였으며, 이러한 과정을 통해 분류된 결과변수의 빈도는 <표 2>와 같다.

**<표 2> 결과변수의 빈도분석 결과**

졸업 후 진로	빈도(%)		total
취업(1)	남	323(34.6)	720(40.6)
	여	397(47.1)	
진학(2)	남	415(44.5)	791(44.6)
	여	376(44.7)	
기타(3)	남	195(20.9)	264(14.9)
	여	69(8.2)	
total	1,775		

### 나. 예측변수

특성화고 학생의 졸업 후 진로결정을 예측하기 위해 패널 데이터의 정보를 최대한 포함하고자 학생 ID를 기준으로 1, 2, 3차년도 데이터를 결합하였으며, 결과변수를 제외하고 가능한 변수를 모두 예측변수로 투입하고 편향성이 최소화된 안정적인 추정 결과를 확보하고자 다음과 같이 데이터를 전처리하였다. 우선 정량화가 어려운 기타 응답, 서술형 응답 등은 결측치로 처리하였고, 전체 데이터의 30% 이상이 결측인 변수의 경우 결측으로 인한 정확도 손실 문제가 나타날 수 있기 때문에(이윤희, 장운선, 2021), 시스템 결측값을 가진 케이스와 무응답 케이스 등이 30% 이상인 변수는 제외하였다. 다만 30% 미만이라 하더라도 결측 비율이 높은 요인들의 경우 의사결정나무 생성에 다수 활용되어 편향성이 높은 결과로 이어질 수 있는 한계가 있기 때문에(남나라, 2021), Breiman(2003)의 제안에 따라 R의 randomForest 패키지에 포함된 na.roughfix 함수를 활용하여 연속 변수는 중앙값으로, 명목변수는 최빈치로 대체하였다. 그 결과 분석에 활용된 예측변수는 1차년도(2016) 259개, 2차년도(2017) 163개로 총 422개이며, 범주별로 구분하면 <표 3>과 같다.



〈표 3〉 예측변수의 범주 및 내용

구분	내용	설명
개인 특성	성별	1=남, 2=여
	거주 지역	1=특별시 2=광역시 3=시 4=읍면
	신장	1=160미만, 2=160이상, 3=170이상, 4=180이상
	체중	1=50미만, 2=50이상, 3=65이상, 4=80이상, 5=100이상
	건강상태	1=매우 허약 ~ 5=매우 건강
	주 평균 운동시간	1=1시간 미만, 2=1시간 이상, 3=2시간 이상, 4= 4시간 이상, 5=10시간 이상
	일 평균 수면시간	1=6시간 미만, 2=6시간 이상, 3=8시간 이상, 4= 10시간 이상
	아침식사	1=먹지 않는다, 2=주 1~2, 3=주 3~4, 4=주 5~6, 5=매일
	행복도	1=전혀 행복하지 않다 ~ 10=매우 행복하다
	고민·걱정거리(8)	(공부/학교/성적, 진학/진로, 가정경제, 부모님과 불화, 애인, 친구, 성격, 외모/신체) 1=없다, 2=보통이다, 3=심각하다, 4=매우 심각하다
	다문화인식(5)	(이웃, 친구, 단짝, 이성친구, 결혼) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	여가시간 활용(11)	(TV 등 영상, 게임, SNS, 인터넷 서핑, 독서, 운동, 동아리/종교활동, 노래방/오락실/만화방, 아르바이트, 집안일, 낮잠/휴식) 1=전혀 하지 않음, 2=월 1회, 3=주 1~2회, 4=주 3~4회, 5=거의 매일
	월평균 독서량***	1=읽지 않음, 2=1권 이상, 3=2권 이상, 4=5권 이상, 5=10권 이상
	독서에 대한 질문(3)	(의무적인 독서, 독서 선호, 책 선물 선호) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
인터넷 중독*(9)	(생활 불규칙, 계획한 일 못함, 다른 일이 있어도 함, 생활이 지루, 초조, 자꾸 생각남, 오랜 시간 함, 사용시간 조절 노력 실패, 주위 사람 지적) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
진로 계획 및 직업 의식	희망 교육수준	1=고등학교, 2=2~3년제 대학, 3=4~6년제 대학교, 4=석사, 5=박사
	미래직업 결정 여부	1=예, 2=아니오
	직업을 가지는 이유	1=생계유지, 2=사회봉사 및 참여, 3=자아실현, 4=사회적 인정
	직업생활 성공 조건	1=돈, 2=노력과 성실성, 3=대인관계, 4=능력, 5=좋은 성품, 6=운/요행, 7=가정배경/인맥, 8=학벌, 9=기타
	인생에서 중요한 요소	1=행복한 가정 이루기, 2=소질과 적성에 맞는 일, 3=돈과 명성, 4=종교생활, 5=사회발전 기여, 6=자기 발전, 7=자녀 교육, 8=건강, 9=여가생활, 10=봉사
	진로에 대한 태도(14)	(새로운 직업 개척, 직업세계 변화 흐름, 직업윤리, 편견/고정관념, 고등교육기관, 대학/전공학과 정보 탐색, 희망직업 정보탐색, 정보 판별, 직업관련 학력/자격, 장기적인 진로계획, 진로계획, 진로준비, 진로선택기중, 어려움 극복) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	자기객관화(5)	(잘하는 일, 좋아하는 일, 삶에서 중요한 것, 결단력, 계획수행) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
가정 배경 및 특성	거주형태	1=부모(보호자)집, 2=하숙, 3=자취, 4=기숙사, 5=친인척 집, 6=내 가정, 7=기타
	월 용돈	1=10만원 미만, 2=10만원 이상, 3=20만원 이상, 4=50만원 이상, 5=100만원 이상
	해외여행경험	1=예, 2=아니오
	전반적 만족도	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	부모님과 활동 빈도(5)	(외식, 영화/연극/뮤지컬, 박물관/미술관/음악회, 여행/등산/낚시, 운동) 1=전혀 하지 않음, 2=연 1~1회, 3=한 학기 1~2회, 4=월 1~2회, 5=주 1~2회

구분	내용	설명	
	부모님과 대화(5)	(학교생활, 흥미/적성, 미래직업, 학교/학과선택, 삶의 가치관) 1=전혀 하지 않음, 2=월 1회, 3=주 1~2회, 4=주 3~4회, 5=거의 매일	
	남성보호자 유형	1=아버지, 2=할아버지, 3=형/오빠, 4=삼촌/외삼촌, 5=고모부/이모부, 6=기타, 7=없음	
	남성보호자 최종학력**	1=무학, 2=초등학교, 3=중학교, 4=고등학교, 5=2~3년제 대학, 6=4년제 대학, 7=석사, 8=박사	
	남성보호자 경제활동**	1=예, 2=아니오	
	남성보호자에 대한 인식* (18)	(방과 후 목적지, 결정에 대한 설명, 남보다 잘함 강조, 사생활 잔소리, 과한 벌, 애정 표현, 생활 파악, 화내는 이유, 높은 기대, 사소한 일에 간섭, 잘못 시 체벌, 함께 시간, 밤 시간, 벌에 대한 합리적 이유, 기대 부담, 개인적인 일 잔소리, 가정폭력, 용기를 줌) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	여성보호자 유형	1=어머니, 2=할머니, 3=언니/누나, 4=고모/이모, 5=숙모/외숙모, 6=기타, 7=없음	
	여성보호자 최종학력**	1=무학, 2=초등학교, 3=중학교, 4=고등학교, 5=2~3년제 대학, 6=4년제 대학, 7=석사, 8=박사	
	여성보호자 경제활동**	1=예, 2=아니오	
	여성보호자에 대한 인식* (18)	(방과 후 목적지, 결정에 대한 설명, 남보다 잘함 강조, 사생활 잔소리, 과한 벌, 애정 표현, 생활 파악, 화내는 이유, 높은 기대, 사소한 일에 간섭, 잘못 시 체벌, 함께 시간, 밤 시간, 벌에 대한 합리적 이유, 기대 부담, 개인적인 일 잔소리, 가정폭력, 용기를 줌) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	부모의 부동산/금융소득**	1=없음, 2=100미만, 3=100~200미만, 4=200~300미만, 5=300~400미만, 6=400~500미만, 7=500~600미만, 8=600~700미만, 9=700~800미만, 10=800~900미만, 11=900~1,000만 미만, 12=1,000만 이상	
	부모의 경제적 책임*	1=고등학교 졸업까지, 2=대학 졸업까지, 3=직장 구할 때까지, 4=결혼까지, 5=평생	
학습 및 성적	중3 성적 등급*	1=1등급 ~ 9=9등급	
	고 2-1 성적*	1=1등급 ~ 9=9등급	
	3-1 과목별 내신등급**(4)	(국어, 수학, 영어, 전문교과) 1=1등급 ~ 9=9등급	
	과목별 흥미*(9)	(국어, 수학, 영어, 과학, 사회, 음악, 미술, 체육, 전문교과) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	주 평균 공부시간	1=주 1시간미만, 2=주 1~5시간미만, 3=주 5~10시간미만, 4=주 10~20시간미만, 5=주 20시간 이상	
	학습방법*(12)	(암기 노력, 기존 지식과 연결, 배울 내용 파악, 끈기있게 공부, 많이 암기, 기존 지식 연결할 때 효율성, 중요한 내용 기억, 최대한 열심히 공부, 반복/밑줄, 기존 지식과 연결방법 고려, 다른 정보 찾기, 계획수행 완료) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	공부이유*(12)	(배움의 즐거움, 나은 사람, 좋은 직업, 돈, 하고싶은 일, 사회에 필요한 사람, 칭찬, 친구들에게 인기, 경쟁상대 이김, 공부 강요, 친구들에게 무시, 벌) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	학업스트레스*(4)	(시험기간 마음 불안, 시험기간 아픈 곳 생김, 시험 종료 후 시험에 대한 걱정 지속, 너무 초조해 생각안 남) 1=전혀 그렇지 않다 ~ 5=매우 그렇다	
	학교 생활	통학 방법*	1=대중교통, 2=자가용, 3=스쿨버스, 4=자전거, 5=도보, 6=택시, 7=기타
		고등학교 선택 이유*	1=대학진학, 2=가정형편, 3=성적, 4=적성, 5=취업, 6=기타



구분	내용	설명
	고등학교 선택 시기*	1=초등학교, 2=중1, 3=중2, 4=중3
	친한 친구 존재 여부*	1=예, 2=아니오
	친구 특성*(4)	(스타일, 관심 분야, 성적, 경제적 수준) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	친구에 대한 생각*(3)	(내 생각에 대한 존중, 고민과 문제 이야기, 외로움) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	전공동아리 경험 여부	1=예, 2=아니오
	기능반 경험 여부	1=예, 2=아니오
	동아리 활동 여부*	1=예, 2=아니오
	현장체험 경험 여부	1=예, 2=아니오
	취업/자격증 방과후 참여	1=예, 2=아니오
	자격증 유무**	1=예, 2=아니오
	자격증 개수**	1=없음, 2=1개, 3=2~4개, 4=5~9개, 5=10개 이상
	교내 수상 여부	1=예, 2=아니오
	진학을 위한 사교육 경험	1=예, 2=아니오
	취업/자격증 사교육 경험	1=예, 2=아니오
재학 중 유급노동 경험	1=예, 2=아니오	
교사 수업 특성	선생님에 대한 질문(4)	(선호, 존경, 이해, 진로/적성 관심) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	담임에 대한 질문(8)	(공정성, 칭찬, 학생 이해, 문제해결 노력, 공부1, 공부2, 상위권 가오, 예습/복습 장려) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	수업만족도(9) (국, 영, 수 각 3문항)	(담당과목 지식수준, 수업역량, 학생이해도 확인) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	전문교과 수업(9)	(수업참여, 수업내용 이해, 실습실/기자재, 전공지식/기술습득, 비 중확대, 현장기술 수준, 선생님의 지식수준, 선생님의 수업역량, 학생이해도 확인) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
학교 특성	전반적 학교만족도	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	학교 시설 만족도(6)	(건물/교실, 운동장/체육시설, 컴퓨터/시청각, 도서실, 학교주변, 교과교실/전공실습실) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	학생들의 수업태도(5)	(집중, 질문, 숙제, 복습, 예습) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	학교 수업 분위기(4)	(잠, 타과목 공부, 잡담, 수업 불참) 1=전혀 그렇지 않다 ~ 5=매우 그렇다
	진로교육 전반적 만족도	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	진로와 직업수업 만족도	1=전혀 만족하지 않다 ~ 5=매우 만족한다
	세부 진로교육 경험(11)	(진로와 직업수업, 진로활동, 진로심리검사, 진로상담, 진로동아 리, 직업인 멘토특강, 현장견학, 학과체험, 현장직업체험, 직업실 무체험, 진로캠프) 1=예, 2=아니오
	취업지도 경험(7)	(취업상담, 채용기업정보제공, 구직활동지도, 취업박람회/채용설 명회, 취업캠프, 졸업생멘토링, 취업처연계) 1=예, 2=아니오
담임 응답	징계 경험*	1=있다, 2=없다
	공부*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	재주*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	리더십*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	꿈*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	가정의 경제적 수준*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	친구관계*	1=전혀 그렇지 않다 ~ 5=매우 그렇다
	학교생활 성실*	1=전혀 그렇지 않다 ~ 5=매우 그렇다

구분	내용	설명
	학업관련 학부모 상담*	1=있다, 2=없다
	생활태도관련 학부모 상담*	1=있다, 2=없다
	진로취업관련 학부모 상담*	1=있다, 2=없다

주1) 변수 내용의 괄호()안의 숫자는 2개 이상인 문항 수임. 그 외는 모두 단일 문항임.

주2) \* 1차년도(2016)에만 조사된 데이터임.

주3) \*\* 2차년도(2017)에만 조사된 데이터임.

주4) \*\*\* 1차년도(2016) 데이터의 경우 결측치 30% 초과하여 삭제함

## 다. 분석방법

본 연구는 한국교육고용패널(KEEP II)의 1, 2, 3차년도 자료를 활용하여 랜덤포레스트(random forest)를 적용함으로써 특성화고 학생들의 졸업 후 진로결정에 영향을 주는 예측요인들을 탐색하고자 한다.

랜덤 포레스트는 Breiman et al.(1984)이 제안한 의사결정나무 기법(CART: classification and regression tree)에 기반 하는 방법으로, 의사결정나무 기법의 경우 판별분석이나 회귀분석에 비해 짧은 수행시간이 소요되며 비교적 간단한 분석 방법과 나무 형태의 도표로 제시되기 때문에 연구자가 분석과정을 직관적으로 이해하고 설명할 수 있다는 장점을 지닌다(유진은, 2015; 최필선, 민인식, 2018). 그러나 설명력은 높지만 예측력이 낮고, 모형이 불안정하며(박창이 외, 2013; 유진은, 2015; Siroky, 2009), 나무의 크기가 커질수록 해당 표본 외의 다른 표본은 충분히 설명하지 못하는 과적합의 문제가 있다(고문정, 2018; Friedman et al., 2000). 이처럼 한 개의 나무로 모형을 훈련시킴으로 인해 발생하는 예측의 불안정성을 극복하기 위해 부트스트랩 샘플링을 통해 다수의 나무를 생성하고 그 결과를 종합하여 예측의 정확성을 높이는 방법들이 제안되었으며, 그 중 Breiman(1996)이 제안한 bagging (bootstrap aggregating)은 부트스트랩 샘플링을 통해 다수의 나무들을 생성하고, 그 결과를 종합하여 모형 전체의 변량은 줄이되 비편향성은 유지하고자 한 앙상블 방식이다. 그러나 bagging은 생성하는 의사결정나무의 개수가 많아져도 최초의 분리에 쓰이는 요인이 비슷할 가능성이 높기 때문에 변량을 효과적으로 조절하는 데 한계가 있다(남나라, 2021).

이러한 점에 기인하여 Breiman(2001)이 제안한 랜덤포레스트 기법은 각 의사결정나무의 성장에 투입하는 예측 요인들을 무작위로 추출된 부트스트랩 표본을 선택하여 활용함으로써 노이즈나 이상치로부터 크게 영향을 받지 않는 강건함을 가지며, 과적합 문제에서 자유로운 안정적인 결과를 얻을 수 있다. 또한 각 의사결정 나무의 설명변수를 채택하는 과정에서도 무작위성을 고려하여 설명변수의 조합을 활용하므로 설명변수가 많은 경우에도 좋은 결과를 보이며, 하나이 강력한 설명변수가 있더라도 각 의사결정 나무가 이에 좌우되지 않고 다양한 설명변수를 고려할 수 있기 때문에 예측의 정확성이 높아지게 된다(김미림, 박민호, 2019 재인용; Dudoit et al., 2002; Hamza, Larocque; 2005). 뿐만 아니라 전통적인 회귀분석이나 로지스틱 회귀분석 모형에서 독립변수의 수가 많아질 경우 독립변수의 외생성이나 오차항 분포에 대한

계약에서 오는 한계점을 보완할 수 있다는 장점이 있다(윤영, 성지훈, 2021). 즉, 종속변수와 예측변수 간의 비선형성 및 예측변수 간의 상호작용을 고려한 추정 결과를 도출할 수 있으며, 다수의 예측변수를 투입하더라도 자유도 감소로 인한 문제로부터 자유롭다는 장점이 있어 가능한 많은 변수들을 모형에 투입할 수 있다(최필선, 민인식, 2018). 따라서 전통적 통계방법의 한계를 보완하는 동시에 다양한 변수를 고려할 수 있다는 점에서 랜덤포레스트 기법은 다른 머신러닝 기법들에 비해 안정적이고 우수한 예측력을 갖는 방법으로 보고되고 있다(김영식, 김훈호, 2019; 박정, 2020; 유진은, 2015; 이동훈, 2019). 이에 개인, 가정, 학교 차원의 여러 변수들 중 특성화고 학생들의 졸업 후 진로결정에 영향을 미치는 주요한 요인이 무엇인지를 탐색하기 위해 랜덤 포레스트 분석이 적합한 연구 방법이라고 판단된다.

본 연구에서는 SPSS 27 프로그램에 R randomForest 확장 패키지를 설치한 후 분석을 실시하였으며, 부트스트랩 표본 수가 많을수록 좋다는 제안(Genuer et al., 2010; Goldstein et al., 2011)에 따라, 의사결정나무 생성 과정을 1000회 반복하도록 설정하였다. 분석 결과는 결과변수에 영향을 미치는 예측변수의 영향력을 크기별로 정리한 중요도 지수(importance index)와 예측변수의 응답에 따른 결과변수의 예측값을 나타내는 부분 의존성 도표(partial dependence plot)를 확인하여 설명변수의 영향력을 파악하였다. 또한 훈련 데이터와 시험 데이터의 정분류율, 정확도, 특이도, 민감도를 확인하여 분류의 정확도 및 모형 과적합성을 판단하였다. 중요도 지수는 모든 나무에서 해당 변수로부터 분할이 일어날 때 불순도의 총 감소량을 나타내는 지수로 각각의 의사결정 나무에 대해 Out-of-bag(OOB) 자료의 오분류율 평균과 해당 변수를 제거한 OOB 자료의 오분류율 평균 차이를 표준오차로 나누어 표준화한 값으로 구해지며, 해당 값이 클수록 순도가 올라가 유의한 변수로 받아들여지므로 중요도가 높은 것으로 해석할 수 있다(김미림, 박민호, 2019). OOB 자료는 훈련 데이터에서 부트스트랩 샘플로 추출되지 않은 약 1/3 상당의 데이터를 가리키는 것으로(남나라, 2021), 대부분의 머신러닝 기법이 훈련 데이터와 시험 데이터로 나누어 분석함으로써 모형 과적합성을 판단하는 반면 랜덤포레스트의 경우 OOB자료를 시험 데이터로 사용할 수 있다는 장점이 있다. 부분 의존성 도표(partial dependence plot)는 결과변수에 대한 다른 변수들의 효과 평균을 제거하고 해당 예측변수의 영향력을 시각화하여 제시하므로, 특정 예측변수의 변화에 따라 결과변수의 변화를 구체적으로 살펴볼 수 있다는 장점이 있다(김진희, 김준엽, 2021; Hastie et al., 2009).

## IV. 연구 결과

### 1. 분석 모형의 예측력

분석 모형의 예측력을 제시하면 <표 4>와 같다. 훈련자료는 전체 자료 중 무선 표집 된 약 2/3의 자료를 지칭하며, 시험 데이터로 사용된 OOB 값은 0.383(38.3%)으로 전체 자료의 약

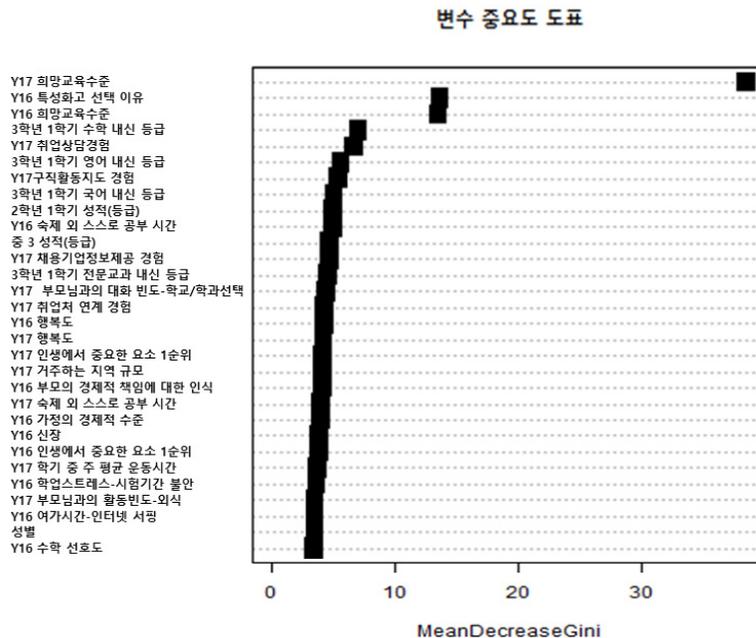
1/3에 해당한다. 머신러닝을 통해 개발한 예측 모델의 경우 훈련자료와 시험자료를 통해 교차 검증이 이루어지며, 이를 통해 도출된 성과 지표를 통해 모형의 예측력을 평가한다(홍기혜, 2020). 랜덤포레스트 분석을 통해 도출된 오차행렬을 통해 예측모형의 정확도, 민감도, 특이도를 계산하였다. 정확도는 72.5%로 취업, 진학, 미결정 각각의 경우를 정확하게 예측한 확률을, 민감도는 72.8%로 실제값이 긍정일 때 예측값 또한 긍정일 확률을, 특이도는 72.2%로 실제값이 부정일 때 예측값 또한 부정일 확률을 의미하며, 랜덤포레스트를 활용한 선행연구와 비교하여 전반적으로 적정 수준임을 확인하였다.

〈표 4〉 분석모형의 예측력

지표	예측력(%)
정확도(accuracy)	72.5
민감도(sensitivity)	72.8
특이도(specificity)	72.2

## 2. 특성화고 학생들의 졸업 후 진로결정 관련 주요 예측변수 탐색

랜덤포레스트 분석 결과 중요도 지수가 높게 나타난 상위 30개 변수는 [그림 1]과 같으며, 중요도 지수가 높은 상위 30개의 예측변수를 영역별로 구분하면 <표 5>와 같다.



[그림 1] 진로결정 관련 중요도 지수 상위 30개 변수

〈표 5〉 진로결정 관련 중요도 지수 상위 30개 변수의 영역별 분류

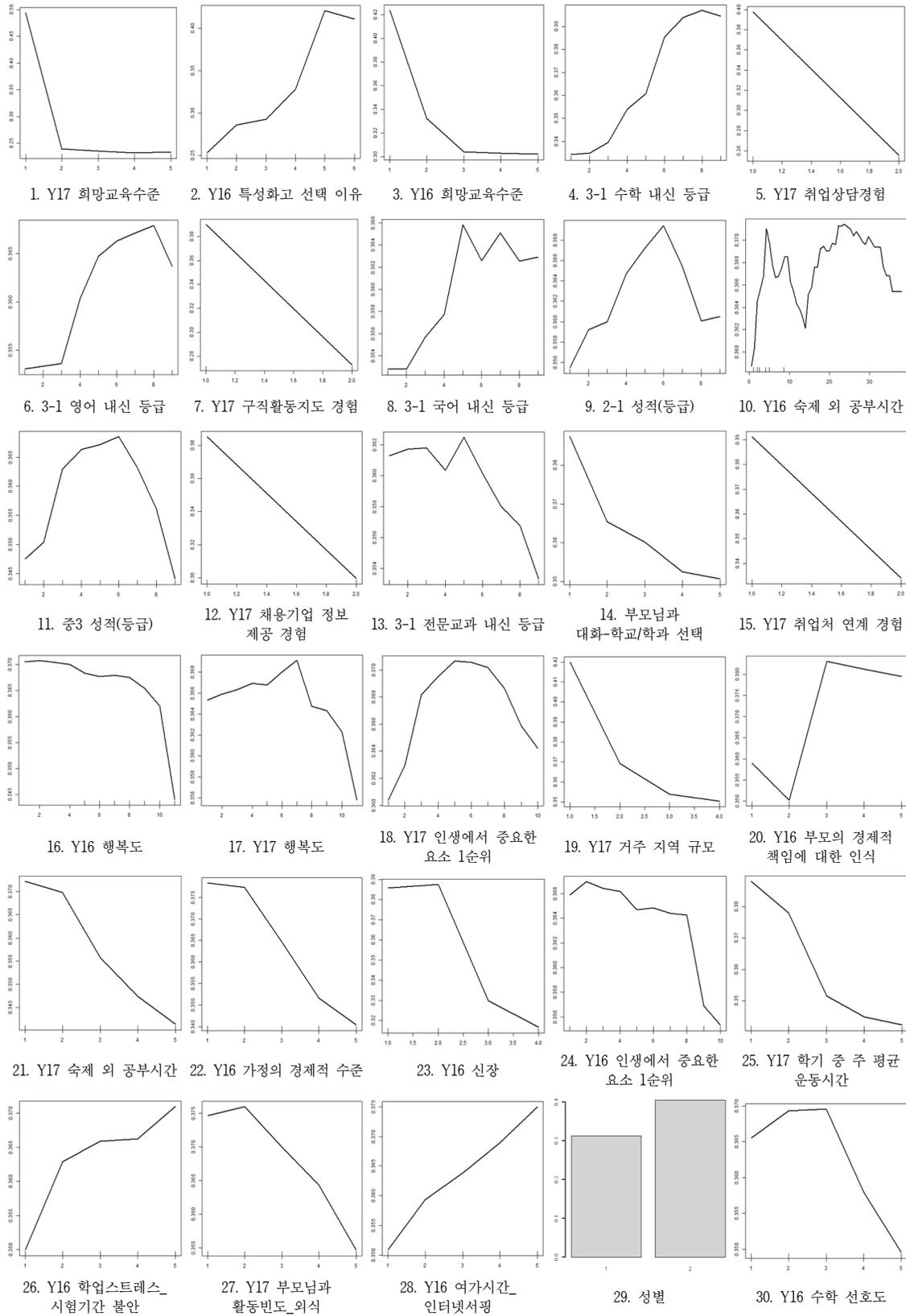
범주	하위범주	변수
개인특성(7)	◦ 일반(3)	성별, Y17 거주 지역 규모, Y16신장
	◦ 여가 및 자기관리(2)	Y17 학기 중 주 평균 운동시간, Y16 인터넷 서핑
	◦ 정의적 특성(2)	Y16 행복도, Y17 행복도
학업 및 성적(10)	◦ 학업(4)	Y16 숙제 외 공부시간, Y17 숙제 외 공부시간, Y16 학업 스트레스_시험기간 불안, Y16 수학 선호도
	◦ 성적(6)	중 3 성적, 2학년 1학기 성적(등급), 3학년 1학기 내신 등급_수학, 영어, 국어, 전문교과
진로 관련 인식(5)	◦ 교육포부(3)	Y16 희망교육수준, Y17 희망교육수준, Y16 특성화고 선택 이유
	◦ 삶의 관점(2)	Y16 인생에서 중요한 요소 1순위, Y17 인생에서 중요한 요소 1순위
가정특성(4)	◦ 일반(2)	Y16 부모의 경제적 책임에 대한 인식 Y16 가정의 경제적 수준
	◦ 부모님과과의 시간(2)	Y17 부모님과과의 대화 빈도 - 학교/학과 선택 Y17 부모님과과의 활동 빈도 - 외식
학교특성(4)	◦ 진로교육 및 활동(4)	Y17 취업상담, Y17 구직활동지도, Y17 채용기업정보제공, Y17 취업처 연계

분석에 사용된 422개 변수 중 가장 중요도가 높은 변수는 고등학교 3학년 시기의 희망교육 수준이었고, 이어서 특성화고 선택 이유, 고등학교 2학년 시기의 희망교육수준, 3학년 1학기 수학 내신 등급, 고등학교 3학년 시기의 취업상담경험으로 나타났다. 그 외 3학년 1학기의 영어, 국어, 전문교과 성적과 2학년 및 중학교 3학년 때의 성적 및 고등학교 2학년과 3학년 시기에 숙제 외에 스스로 공부하는 시간도 진로결정을 예측하는 상위 변인으로 분석되었다. 학업 및 성적관련 변인 외에 부모님과과의 대화 및 가정의 경제·문화적 배경과 학교의 진로교육 활동 및 자신의 신장, 행복도, 삶에 대한 관점 등이 특성화고 학생들의 졸업 후 진로결정에 영향을 주는 상위 예측요인으로 나타났다. 영역별로 살펴보면, 상위 30개 예측변수 중 학업 및 성적 관련 변수가 총 10개로 가장 많이 포함되었으며, 다음으로 개인특성(7), 진로 관련 인식(5), 그리고 가정특성과 학교특성 관련 변수가 각각 4개씩 포함되었다. 특히 학업, 성적, 진로에 대한 인식 및 개인특성 변수 등 22개 변수가 학생 개인 요인으로 분석되었는데, 이러한 결과는 특성화고 학생들의 졸업 후 진로결정 예측에 있어 가정 및 학교관련 요인보다 개인요인의 중요성을 시사한다고 할 수 있다.

### 3. 진로결정과 주요 예측변수와의 관계

본 절에서는 특성화고 학생들의 졸업 후 진로결정 예측력이 높은 상위 30개 변수의 부분의존성 도표를 통해 결과변수와 예측변수 간의 관계를 확인하고자 한다[그림 2]. 부분의존성 도표는 결과변수에 대한 예측 결과의 방향성이나 증감의 경향을 살펴볼 수 있으나, 예측변인과 종속변인의 인과관계를 나타내는 것은 아니다(김길재 외, 2023). 도표의 X축은 각 예측변수의 범위를, Y축은 결과변수에 미치는 예측변수의 효과 범위를 즉, 취업, 진학, 미결정의 부분의존성을 의미한다.

중요도 지수가 높은 순서로 유사한 범주의 문항들을 살펴보면, 먼저 고2(Y16) 시기와 고3(Y17) 시기의 희망교육수준이 높다고 응답한 학생들의 취업 선택 예측력이 높게 나타났고, 3학년 성적의 경우 수학, 영어, 국어 내신 등급과 전문교과의 내신 등급은 진로결정의 패턴이 반대의 양상을 보였다. 예컨대 구체적인 양상은 다소 다르게 나타났으나 수학, 영어, 국어 교과의 경우 3학년 1학기 내신 등급이 높을수록 취업을 선택한 확률이 높았으며, 전문교과의 경우 중간이하 등급 학생들의 취업 선택 예측률이 높게 나타났다. 그러나 고등학교 2학년 성적과 중학교 3학년 시기의 경우 성적이 아주 높거나 아주 낮았던 학생들이 졸업 후 취업을 선택하는 양상을 보였으며, 고등학교 2학년 때의 수학선도도가 보통 이하로 낮은 학생들의 진로 미결정 확률이 높게 나타난 것을 확인할 수 있다. 또한 대체로 숙제 외 공부시간이 많다고 응답한 학생들의 취업 결정 예측률이 높은 것으로 나타났다. 즉, 성적관리 및 스스로 학습하는 태도가 높은 학생들의 취업 결정 예측률이 높으며, 이는 중학교의 기초학업능력이 낮은 학생들이 특성화고등학교에 진학하여 학업에 대한 관심이 적고, 일반고 학생에 비해 생활 태도 측면의 이탈 행동의 비율이 높다는 사회적 편견(김정숙, 나민주, 2015; 이휘인, 김현철, 2015)을 줄이기 위한 논의가 필요함을 시사한다.



[그림 2] 진로결정 상위 30개 예측변수의 부분의존성 도표

한편, 고등학교 3학년 시기에 경험한 진로교육 및 활동 중 취업상담, 구직활동지도, 채용기업 정보 제공, 취업처 연계 활동이 높은 순위로 예측되었으나, 각 활동 경험이 없다고 응답한 경우의 취업 선택 예측력이 더 높은 것으로 나타나 학교에서 이루어지는 관련 활동 경험 그 자체보다 개별 활동에 대한 구체적인 만족도 조사를 통해 학생 맞춤형 진로교육과 진로활동으로 내실화 할 필요가 있다고 판단된다.

그 외 유사한 경향을 보인 예측변수를 살펴보면, 부모님과 대화\_학교/학과 선택(Y17), 거주지역 규모(Y16), 숙제 외 공부시간(Y17), 가정의 경제적 수준(Y16), 키(Y16), 주 평균 운동시간(Y17), 부모님과 활동\_외식(Y17)은 각 변수의 값이 클수록 취업을 선택할 가능성이 점차 높아지는 예측률을 보였으며, 학업스트레스 중 시험기간 불안(Y16)이 낮고, 여가시간에 인터넷 서핑(Y16)을 즐기는 빈도가 적은 학생들의 취업 결정 예측률은 상대적으로 높은 양상을 보였다. 행복도의 경우 고2 시기(Y16)와 고3 시기(Y17)의 구체적인 양상은 다소 다르게 나타났으나 매우 행복하다(10)고 느끼는 학생이 취업을 선택할 것으로 예측될 가능성이 높은 것을 확인할 수 있으며, 남학생에 비해 여학생의 졸업 후 진로 미결정 확률이 높게 나타났다. 한편, 고등학교 2학년 때 ‘소질과 적성에 맞는 일’이 인생에서 가장 중요한 요소라고 생각하고, 그 외 ‘좋은 배우자와 행복한 가정’, ‘돈과 명성’, ‘종교생활’, ‘사회발전에 기여’, ‘자기 발전’, ‘자녀 교육’, ‘건강한 삶’이라고 생각하는 학생들의 경우 취업보다는 진학 또는 진로 미결정을 할 확률이 높게 나타났으나, 3학년 시기에는 ‘사회발전에 기여’를 가장 높게 인식한 학생들의 졸업 후 진로결정에 대한 고민이 많음을 확인할 수 있다.

## V. 요약 및 결론

본 연구는 특성화고등학교 학생들의 졸업 후 진로 결정에 미치는 주요한 영향요인이 무엇인지를 종단적 관점에서 탐색하여 학생들의 진로결정에 도움을 줄 수 있는 방안 마련 및 현장에서의 교육적 활용성을 높이고자 하였다. 이를 위해 한국교육고용패널 II(Korean Education & Employment Panel: KEEP II)의 1차, 2차, 3차년도 자료를 활용하였으며, 주요 분석결과를 바탕으로 한 결론 및 시사점은 다음과 같다.

먼저 이론적 근거 설정 후 가설 검증의 단계로 진행되는 기존 연구의 경우 다중공선성 문제 등으로 인해 투입할 수 있는 변수의 수가 제한적이기 때문에, 문헌탐색을 통해 결과변수와 관련이 있을 것으로 예상되는 다수의 변인이 있다 하더라도 전체 중 일부만을 이용하여 모형을 구축하는 것이 일반적이다(유진은, 2015). 또한 관측 변수의 평균값을 기반으로 잠재 변수를 활용한 분석이 이루어지기 때문에 각각의 관측 변수가 가진 정보를 해석하는 데에는 어려움이 있다는 제한점이 있다. 따라서 본 연구는 특성화고 학생들의 졸업 후 진로 결정 예측 요인을 검증하기 위한 이론적 근거를 제시했을 뿐만 아니라, 결과변수와 예측변수 유형의 제약조건이 없고(조용준, 2018), 다수의 예측변수를 투입하더라도 자유도 감소로 인한 문제로부터 자유롭

다(최필선, 민인식, 2018)라는 랜덤포레스트 분석 방법의 장점을 활용하여 예측 요인을 분석하고 각각의 요인에 대한 중요도를 도출했다는 점에서 기존 연구와 차별성이 있다. 또한 취업, 진학, 미결정 각각의 경우를 정확하게 예측한 정확도는 72.5%, 실제 취업을 선택한 학생을 바르게 예측하는 민감도는 72.8%, 그 반대의 경우인 특이도는 72.2%로 기존의 선행연구와 비교했을 때 예측 성능이 우수하다. 이러한 점에서 랜덤포레스트를 활용한 본 연구의 접근은 특성화고등학교 학생들의 졸업 후 진로결정에 영향을 미치는 다양한 교육 맥락적 변인들을 종합적으로 탐색했다는 점에서 의의가 있으며, 특성화고등학교가 특정분야 인재 및 전문 직업인 양성을 위한 취업중심·산학협력 학교라는 본래의 설립 목적에 맞게 운영되고, 그 과정에서 학생들이 올바른 자아관과 미래관을 바탕으로 한 진로결정에 이르는 교육적 효과성을 극대화할 수 있는 기초자료를 제공할 수 있다는 시사점이 있다.

둘째, 특성화고 학생들의 졸업 후 진로결정에 영향을 주는 요인으로 ‘고등학교 3학년 시기의 희망 교육수준’이 가장 높은 예측변인으로 나타났으며, 다음으로 ‘특성화고 선택 이유’, ‘고등학교 2학년 시기의 희망교육수준’, ‘3학년 1학기 수학 내신 등급’, ‘3학년 때 경험한 취업상담’의 순으로 중요도가 높은 것으로 나타났다. 구체적인 양상은 다소 다르게 나타났으나 수학, 영어, 국어 교과목의 경우 3학년 1학기 내신 등급이 높을수록 취업을 선택한 확률이 높았으며, 전문교과목의 경우 중간이하 등급 학생들의 취업 선택 예측률이 높게 나타났다. 고등학교 2학년과 중학교 3학년 성적도 높은 순위의 예측변인에 포함되었는데, 두 시점 모두 성적이 아주 높거나 아주 낮았던 학생들의 취업 예측률이 높으며, 고등학교 2학년 때의 수학선호도가 낮을 경우 취업을 결정할 확률이 높은 것을 확인할 수 있었다. 또한 대체로 숙제 외 공부시간이 많다고 응답한 학생들의 취업 결정 예측률이 높은 것으로 나타났다. 즉, 성적관리 및 스스로 학습하는 태도가 높은 학생들의 취업 결정 예측률이 높으며, 이는 중학교의 기초학업능력이 낮은 학생들이 특성화고등학교에 진학하여 학업에 대한 관심이 적고, 일반고 학생에 비해 생활 태도 측면의 일탈 행동의 비율이 높다는 것(김정숙, 나민주, 2015; 이휘인, 김현철, 2015)이 사회적 편견이며 일반화의 오류일 수 있다는 점과 특정분야 인재와 전문 직업인 양성을 위한 특성화고등학교의 역할에 대한 사회적 논의가 필요함을 시사한다. 한편, 고등학교 3학년 시기에 경험한 진로교육 및 활동 중 취업상담, 구직활동지도, 채용기업 정보 제공, 취업처 연계 활동이 높은 순위로 예측되었으나, 각 활동 경험이 없다고 응답한 경우의 취업 선택 예측력이 더 높은 것으로 나타나 학교에서 이루어지는 관련 활동 경험 그 자체보다 개별 활동에 대한 구체적인 만족도 조사를 통해 학생 맞춤형 진로교육과 진로활동으로 내실화 할 필요가 있다고 판단된다.

셋째, 분석에 투입한 422개 변수 중 중요도 지수 상위 30개 변수에 학업 및 성적 관련 변수가 총 10개로 학생들의 졸업 후 진로결정을 예측하는 중요한 변수로 고려되었으며, 성별, 키, 지역 규모, 여가시간, 행복도 등 개인특성 관련 변수(7개)와 교육포부와 삶의 관점 등 진로에 대한 인식 관련 변수(5개)도 주요한 예측 변수로 도출되었다. 특히 가정 특성 변수(4)와 학교 특성 변수(4)에 비해 학생 개인 요인이 다수를 차지한 이러한 결과는 특성화고 학생들의 졸업 후 진로결정을 예측하는 데 있어서 학업 및 진로와 관련한 개인적 요소와 태도가 무엇보다 중

요한 역할을 한다는 점을 시사한다. 선행연구에서도 성별, 학업성취, 자기이해, 적성, 가정만족도, 부모와의 대화, 학교만족도, 진로교과와 진로체험 등이 특성화고 학생들의 진로결정에 영향을 미치는 요소라고 밝혔다(김지효, 2018; 신희경, 김우영, 2005; 오석영, 2012; 오현석 외, 2010). 그러나 본 연구의 결과에서 도출된 중요도 지수와, 예측변수와 진로결정과의 관련성을 참고하여 특성화고등학교의 교육과정 및 진로교육 설계 시 어떠한 변수를 우선적으로 고려할지에 대한 논의가 가능하며, 선행연구에서 언급되지 않은 학생의 행복도와 외모(신장), 삶의 우선순위 등도 고려될 필요가 있을 것이다.

본 연구의 의의에도 불구하고 후속 연구에서 다루어야 할 몇 가지의 제한점은 다음과 같다. 첫째, 본 연구에서는 취업, 진학, 미결정 집단을 구분하기 위해 특성화고 패널 데이터를 대상으로 3차년도 현재학력에 대한 응답이 대학 재학 중인 경우 '진학'으로 분류하였으며, 고등학교가 최종학력인 학생 데이터 중 현재 일자리가 있는 학생은 '취업'으로, 그렇지 않을 경우 '기타'로 분류하였고, 고교유형과 졸업 후 진로가 확인되지 않는 학생 데이터는 제외하였다. 그러나 3차년도 데이터는 2018년 현재 고3 학생들의 졸업 후 1년이 지난 시점인 2019년에 조사된 자료로, 학교 졸업과 동시에 이루어진 진로 결정이라고 판단하기에는 취업 후 퇴직 등 몇 가지 다른 변수가 발생할 수 있는 가능성이 있다. 따라서 추후 조사에서는 졸업 후 진로결정의 변화를 직관적으로 확인할 수 있는 학생 응답 변인이 추가될 필요가 있으며, 이를 통해 추후 연구에서는 보다 세분화되고 다양한 방법으로 진로결정 집단을 구분하여 영향 변인을 탐색하고 기존 연구와 비교해 볼 필요가 있을 것이다. 둘째, 본 연구는 특성화고 졸업생의 진로결정을 설명하는 다양한 맥락적 변인을 탐색하는 것을 목적으로 하였으나, 각각의 변인에 대한 깊이 있는 분석이 이루어지지 않았다. 따라서 본 연구의 결과로 도출된 중요도 지수가 높은 변인들에 대한 추가적인 연구 수행을 통해 특성화고 학생들의 진로결정에 도움을 줄 수 있는 구체적인 교육적 활용 방안을 마련할 수 있을 것이다.

## ❖ 참고문헌 ❖

- 강민수(2021). 특성화고 학생들의 진로결정 네트워크 분석. 진로교육연구, 34(1), 1-23.
- 고문정(2018). 누가 온라인 뉴스에 댓글을 작성하거나 뉴스를 공유하는가?: 통계 학습 방법의 적용. 사이버커뮤니케이션학보, 35(1), 5-51.
- 고용노동부 (2017). 「2016~2026년 중장기 인력 수급 전망」. 세종: 고용노동부.
- 교육부(2019) “청년들의 성장경로 다양화”를 위한 「고졸취업 활성화 방안」 발표. 세종: 교육부.
- 길혜지, 백순근, 양현경(2018). 마이스터고 학생과 특성화고 학생의 진로준비 경험과 진로성숙도 비교 분석. 직업능력개발연구, 21(3), 225-248.
- 김경근, 변수용(2006). 한국 사회에서의 상급학교 진학 선택 결정요인. 교육사회학연구, 16(4), 1-27.
- 김경년(2011). 고교유형의 다양화를 통한 교육적 선발이 학생들의 직업포부 형성에 미치는 영향. 교육학연구, 49(4), 121-145.
- 김기현, 방하남(2004). 고등 교육 진학에 있어 가족배경의 영향과 성별 격차: 한국과 일본의 경우. 한국사회학, 39(5), 119-151.
- 김길재, 박강윤, 문명현(2023). 랜덤포레스트를 활용한 청소년의 비행 예측 요인 탐색. 청소년학연구, 30(5), 267-290.
- 김미림, 박민호(2019). 랜덤포레스트를 활용한 대학생의 최초 취업 사교육 참여 시점별 특성 분석. 교육연구논총, 40(1), 1-33.
- 김성환, 전용성(2004). 청소년 진로선택결정 요인 - 가정환경을 중심으로. 제1회 한국교육고영패널 학술대회 자료집. 245-258, 한국직업능력개발원.
- 김수정, 송성화, 조민정 (2013). 전문계 고등학교 학생의 진로준비행동에 영향을 미치는 변인 연구. 진로교육연구, 26(4), 85-107.
- 김영식, 김훈호(2019). 머신러닝 기법을 활용한 사교육 참여 예측 모형 탐색. 교육재정경제연구, 28(3), 29-52.
- 김영식, 이호준(2021). 고등학생의 학교 참여 예측 요인 탐색: 학생회 활동 및 학교 경영 참여를 중심으로. 교육연구논총, 42(1), 53-78.
- 김정란(2006). 실업계고등학교 학생의 대학진학 결정요인에 관한 연구. 경희대학교 대학원 석사학위 논문.
- 김정숙, 나민주(2015). 선취업-후진학 체제 이후 특성화고 학생들의 변화 양상: A학교 사례연구. 교원교육, 31(3), 219-245.
- 김지효(2018). 특성화고 졸업생의 취업에 미치는 영향요인 분석. 취업진로연구, 8(3), 1-20.
- 김진원, 모영민(2019). 특성화고와 마이스터고 학생의 진로성숙도 및 영향요인 비교 분석 연구. 직업능력개발연구, 22(1), 37-73.
- 김진희, 김준업(2021). 랜덤 포레스트 회귀를 이용한 중·고등학교 1학년의 교우관계 영향요인 탐색. 교육방법연구, 33(3), 589-611.
- 김흥기, 김지훈, 김양현(2017). 왜 특성화고 학생들은 대학에 진학하고자 하는가?: 특성화고 학생의 진로 수정 경험에 관한 질적 연구. 한국사회, 18(2), 35-69.

- 남나라(2021). 혼합효과 랜덤 포레스트 기법을 활용한 중학생의 학업성취도 예측 요인 분석. 서울대학교 대학원 박사학위논문.
- 노경란, 변정현 (2016). 특성화고 졸업자의 입직 및 초기직장적응에 대한 질적 연구. 직업교육연구, 35(1), 21-43.
- 두민영, 우버들, 신창호 (2018). 능력중심사회에 대한 특성화고 교사들의 인식 분석. 교육문화연구, 24(2), 419-438.
- 박윤희, 허영준, 오계택 (2013). 고졸자 채용에 대한 특성화고 교사, 학생, 학부모의 인식 비교 연구. 직업교육연구, 32(3), 117-140.
- 박정(2020). 머신러닝을 활용한 서울시 중학생 진로성숙도 예측 요인 탐색. 한국빅데이터학회지, 5(2), 155-170.
- 박창이, 김용대, 김진석, 송종우, 최호식(2013). R을 이용한 데이터마이닝 (개정판). 서울: 교우사.
- 배정수, 송병국(2015). 의사결정나무 분석기법을 활용한 고등학생 진로결정수준 결정요인 우선순위 탐구. 진로교육연구, 28(4), 79-105.
- 신희경, 김우영(2005). 우리나라 고등학생의 진로결정수준과 결정 동기에 관한 연구: 개인, 가정, 학교의 역할. 직업능력개발연구, 8(2), 47-83.
- 안선영·이경상(2009). 중등교육단계 청소년 진로결정요인 및 결과: 교등학교 계열선택을 중심으로, 미래청소년학회지, 6(4): 81-99.
- 오석영(2012). 특성화고 학생의 진로결정요인 분석: 서울지역 진로결정 집단별 비교. 직업교육연구, 31(3), 135-151.
- 오현석, 배진현, 이상은, 장현진(2010). 고등학생의 진로결정 수준 비교연구. 직업교육연구, 29(2), 245-265.
- 유진은(2015). 랜덤포레스트: 의사결정나무의 대안으로서의 데이터마이닝 기법. 교육평가연구, 28(2), 427-448.
- 윤영, 성지훈(2021). 고등학생의 성별에 따른 공동체 역량 예측변인 탐색: 랜덤 포레스트를 활용하여. 청소년학연구, 28(4), 383-410.
- 윤초희, 안태희(2015). 특성화고등학교 학생의 학교선택 동기, 진로신념과 지각된 사회적 지지가 진로준비행동에 미치는 영향. 청소년학연구, 22(8), 137-157.
- 이동훈(2019). 기계학습 분류 기법을 활용한 대졸자 취업 예측 모델 연구. 단국대학교 대학원 석사학위논문.
- 이명훈(2015). 특성화고 1학년 학생들의 진로결정과 관련 변인. 한국기술교육학회지, 15(1), 68-86.
- 이명훈(2016). 특성화고 학생들의 진로결정과 관련 변인간의 구조적 관계 분석. 한국기술교육학회지, 16(2), 149-166.
- 이병욱, 안재영, 이찬주, 이상현(2013). 특성화고 진학 및 졸업 후 진로에 대한 중학생, 학부모, 산업체 인사 담당자의 인식 비교 연구. 대한공업교육학회지, 38(2), 48-67.
- 이은경(2014). 특성화고등학교 학생들의 진로선택에 영향을 미치는 요인에 관한 연구-상업계열을 중심으로. 경희대학교 교육대학원 석사학위논문.
- 이정민, 정혜원(2019). 특성화 및 마이스터고 학생의 진로성숙도에 관한 학생 및 학교요인 분석. 한국청소년연구, 30(1), pp 279-311.
- 이지혜, 정철영(2010). 특성화고 학생의 취업 장벽과 관련 변인. 농업교육과 인적자원개발, 42(1), 25-47.

- 이휘인, 김현철(2015). 특성화고 학생의 진로결정 및 중학생의 특성화고 선택에 영향을 미치는 요인 탐색-2014 서울교육중단연구를 중심으로. *교육문화연구*, 21(6), 169-190.
- 조균제, 김선진(2009). 고등학교 상업계 학생의 진로의식을 통한 진로지도의 개선방안에 관한 연구. *경영교육연구*, 54, 327-348.
- 조용, 임언, 이강주, 이호영, 박최훈, 박보경(2011). 특성화고 진로이력분석 연구 2011. 서울특별시 교육청.
- 조용준(2018). 빅데이터 SPSS 최신 분석기법. 서울: 한나래.
- 전현정, 정혜원(2019). 직업계고 학생의 진로결정 및 진로개발효능감의 분석: 2006년과 2016년 비교를 중심으로. *교육학연구*, 57(1), 49-80.
- 정승(2020). 특성화고 졸업생의 취업 후 현황 및 직장 만족도에 미치는 영향요인 검증. *서울교육 이슈페이퍼*, 31(7), 1-10.
- 정윤경(2000). 진로상담을 위한 고등학생의 진로결정문제 유형에 관한 연구. 한국직업능력개발원, 연구노트 00-5.
- 최동선(2008). 전문계 고교생의 대학 진학 계획: 1차년도와 4차년도의 비교. 제4회 한국교육 고용 패널 학술대회자료집.
- 최보금(2020). 고등학생 진로성숙도에 영향을 미치는 학생 및 학교 변인 탐색: 고교유 형의 영향력을 중심으로. *진로교육연구*, 33(2), 49-66.
- 최필선, 민인식(2018). 머신러닝기법을 이용한 대졸자취업 예측모형. *직업능력개발연구*, 21(1), 31-54.
- 채창균(2009). 전문계 고교 졸업생의 진로 선택. *HRD review*, Spring, 59-74.
- 홍기혜(2020). 랜덤포레스트 머신러닝 알고리즘 기반 남·여 청소년의 자살생각 예측 및 분석. *한국사회복지학*, 72(3), 157-180. <http://doi.org/10.20970/kasw.2020.72.3.007>
- 홍성표, 정진철(2016). 특성화고등학교 학생의 취업·진학결정수준과 준비행동에 따른 진로결정 유형화 및 영향요인 분석. *직업교육연구*, 35(5), 1-30.
- Breiman, L.(1996). Bagging predictors. *Machine Learning*, 24, 123 - 140. <https://doi.org/10.1007/BF00058655>
- Breiman, L.(2001). Random forests. *Machine Learning*, 45, 5 - 32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J.(1984). *Classification and regression trees*. Belmont: Wadsworth.
- Coleman, J. S.(1990). *Equality and Achievement in Education*. Boulder, CO: Westview Press.
- Crites, J.O.(1981). *Career counseling: Models, methods, and materials*. New York: McGraw-Hill.
- Dudoit, S., Fridlyand, J., & Speed, T. P.(2002). Comparison of discrimination methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, 97, 77-87.
- Friedman, J., Hastie, T., and Tibsirani, R.(2000). Additive logistic regression: a statistical view of boosting. *The annals of statistics*, 28(2), 337-407.
- Gati, I.(1986). Making career decisions- Asequential elimination approach. *Journal of Counseling Psychology*, 33: 408-417.

- Genuer, R., Poggi, J.-M., & Tuleau-Malot, C.(2010). Variable selection using random forests. *Pattern Recognition Letters*, 31, 2225 - 2236.
- Goldstein, B. A., Polley, E. C., & Briggs, F.(2011). Random forests for genetic association studies. *Statistical Applications in Genetics and Molecular Biology*, 10(1), 32.
- Gordon, V.N.(1981), The Undecided Student: A Developmental Perspective. *The Personnel and Guidance Journal*, 59: 433-439.
- Hamza, M., & Larocque, D. (2005). An empirical comparison of ensemble methods based on classification trees. *Journal of Statistical Computation and Simulation*, 75(8), 629-643.
- Hastie, T., Tibshirani, R., & Friedman, J. H.(2009). *The elements of statistical learning: data mining, inference, and prediction(2nd ed.)*. New York: Springer.
- Krumboltz, J. D., & Drumboltz, H. B. (1976). *Changing children's behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Liaw, A., & Wiener, M. (2018). *RandomForest: Breiman and Culter's random forests for classification and regression*. R package(version 4.6-14)
- Mitchell, K.E., Levin, A.S., & Krumboltz, J.D.(1999). Planned happenstance: Constructing unexpected career opportunities. *Journal of Counseling and Development*, 77: 115-124.
- Parsons, F.(1909). *Choosing a vocation*. Boston: Houghtom Mifflin.
- Siroky, D. S. (2009). Navigating random forests and related advances in algorithmic modeling. *Statistics Surveys*, 3, 147-163.
- KOSIS 국가통계포털 <https://kosis.kr/search/search.do?query=%EC%8B%A4%EC%97%85%EB%A5%A0>

## ❖ Abstract ❖

### Exploration of Predictive Factors on Career Decisions of vocational high school students using Random Forests

Kim Ha-Jeong, Won Hyo-Heon(Pukyong National University)

The purpose of this study is to explore major variables that affect students' career decisions after graduation from vocational high schools Using data from the first, second, and third years of the Korean Education & Employment Panel(KEEP II) and random forest analysis.

As a result, it was found that 'desired education level' in the third year of high school was the most important factor, followed by 'reason for choosing a vocational high school', 'desired education level' in the second year of high school, 'math grade in the first semester of the third year', and 'career counseling' experienced in the third year. Academics and grades-related variables, which included 10 of the top 30 variables in the importance index, were considered important factors in predicting students' career decisions after graduation. In addition, personal characteristics such as gender, height, the size of region, leisure time, and happiness (7 variables), career perception such as educational aspirations and perspectives in life (5 variables), family characteristics (4 variables), and school characteristics (4 variables) were also derived as major Predictive Factors. Based on the results of this study, educational implications for preparing educational plans that can help graduates of vocational high schools in their career decisions were discussed.

**Key words:** career decision, vocational high school, career guidance, random forests